Title
Implicit Communication Through Behavioral Policy Design

Author
Reiff, Joseph

Publication Date
2023

UNIVERSITY OF CALIFORNIA

Los Angeles

Implicit Communication Through Behavioral Policy Design

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy in Management

by

Joseph Summer Reiff

2023

ABSTRACT OF THE DISSERTATION

Implicit Communication Through Behavioral Policy Design

by

Joseph Summer Reiff

Doctor of Philosophy in Management

University of California, Los Angeles, 2023

Professor Hal E. Hershfield, Chair

Policymakers are increasingly using interventions, informed by psychological science, to motivate behavior change ("behavioral policies") across a range of domains: from reducing failure to appear in court to increasing COVID19 vaccinations. Field experiments testing these behavioral policies are theoretically and practically important: they provide ecological validity to psychological theories, while evaluating real solutions to pressing social problems. Yet, recent evidence suggests, when tested in interventions in the field, promising ideas from psychology often yield null effects or even backfire. Moreover, behavioral policies testing the same psychological principle often result in divergent results across field experiments. This presents multiple obstacles for behavioral science. Without theories that can explain unexpected and divergent evidence across field experiments, researchers may prematurely conclude that psychological principles that have been theorized to be highly motivating do not have ecological validity, and policymakers may prematurely conclude that behavioral policies are not useful tools for addressing problems at scale. While recent evidence has pointed to differences in sampling and measurement to explain unexpected and divergent results, my dissertation explores a different hypothesis: people draw inferences about policymakers from details in the design of behavioral policies and the contexts in which they are implemented, which can help determine when policies work as intended and when they backfire. My dissertation includes four chapters, each highlighting an inference people draw from behavioral policies: inferences about urgency, support, inauthenticity, and trustworthiness. By identifying social inferences as a hidden mechanisms underlying the effects of behavioral policies, my research provides new insight into (1) the unexpected effects of behavioral policies that both academics and expert practitioners previously believed to be effective, (2) the redesign of psychological interventions so that they have their intended positive effects, (3) the identification of moderators that help explain the divergent effects of interventions across contexts, and (4) the inclusion of consequential yet unexplored outcomes in the evaluation of behavioral policies. Together, my dissertation integrates and advances knowledge on social inference making and behavior change, provides a new perspective on studying psychological theories in the field, and offers practical implications for policymakers seeking to design effective behavioral policy.

The dissertation of Joseph Summer Reiff is approved.

Hengchen Dai

Jana Gallus

Katherine Milkman

Stephen A. Spiller

Hal E. Hershfield, Committee Chair

University of California, Los Angeles

2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

At UCLA's admit day in Spring 2017, my mind was already made up; I had decided to join the inaugural class of the UCLA Behavioral Decision Making PhD program. My decision was based on three factors: (1) the faculty at UCLA seemed to be deeply invested in the students' success, (2) the group was focused on research with real-world impact, and (3) the other admitted students—namely, David Zimmerman and Jon Bogard—seemed to share my aspirations to conduct good, impactful research. Outcome bias aside, I now can confidently say I made an excellent decision.

I am extremely grateful to my advisor and mentor, Hal Hershfield, for his unwavering dedication to my success as a researcher and aspiring academic. He taught me how to write clear papers (over many agonizing hours at Zinqué), create engaging presentations (going through every slide with me before my first BDM lab), and evaluate others' research (dissecting the pros and cons of presentations we watched together). Hal's kindness and humor made graduate school a more enjoyable, less stressful experience.

I am indebted to Hengchen Dai and Jana Gallus for inviting me to join *the team*, where I found my passion conducting field research. I am so thankful to have had such conscientious and tenacious mentors. They both spent so many hours in the weeds with me, teaching me the "nitty gritty" of field research. Hengchen gave me confidence in myself in graduate school when nothing seemed to be going my way. She is incredibly thoughtful, always considering what would put me in the best possible position. Jana brings such joy to every discussion about research, radiating an intrinsic passion for learning about human motivation. Her positive outlook is inspiring and hopefully contagious.

Stephen Spiller provided selfless support throughout graduate school, helping with stats questions on nearly all my projects. Stephen's commitment to improving science affected the norms and expectations within our group, and I am deeply grateful for that. Katy Milkman provided exceptional guidance and feedback on our project together. Getting to work with one of my academic role models remains a highlight of graduate school. Her leadership and innovation in the field has been deeply influential to my aspirations. I am also grateful for the rest of the BDM faculty, in particular to those I had the pleasure to work with over the years including Cassie Mogilner Holmes, Eugene Caruso, Kareem Haggag, and Craig Fox.

I am so lucky to have joined the program with David and Jon (and our behavioral marketing cousin, Ipek). I could not have asked for a better cohort. Thank you to the rest of the BDMers (Yilin, Malena, Ilana, Megan, Jo, and Kate) for making graduate school surprisingly fun.

I am grateful to my loving and supportive family—Caryn, Brad, Nathan, Garret, David, and Zona—for the thousands of calls in which you listened to me kvetch, and the wonderful vacations we had together throughout the last six years that gave me perspective when graduate school felt all consuming. I am so lucky to have family that doubles as a group of best friends. I pinch myself every morning.

Lastly, one thing I did not realize on that UCLA admit day in 2017 was that I had also met my life partner, Julia. Her support and care throughout the highs and lows of graduate school mean everything to me. I learn from her everyday how to be present and compassionate. She makes me a better person. I cannot wait for our future together in Maryland and beyond.

Thank you.

Chapter 1 is a reprint of Reiff, Joseph S., Hengchen Dai, John Beshears, Katherine L. Milkman, and Shlomo Benartzi (2023), "Save More Today or Tomorrow: The Role of Urgency in Pre-commitment Design." Forthcoming, *Journal of Marketing Research*. The initial field experiment, post-test, and prediction surveys were collected by Hengchen Dai, John Beshears, Katherine L. Milkman, and Shlomo Benartzi before I joined the project. The remaining empirical work, theorizing, and writing was joint with Hengchen Dai, John Beshears, Katherine L. Milkman, and Shlomo Benartzi.

Chapter 2 is a reprint of Reiff, Joseph S., Justin C. Zhang, Jana Gallus, Hengchen Dai, Nathaniel M. Pedley, Sitaram Vangala, Richard K. Leuchter, Gregory Goshgarian, Craig R. Fox, Maria Han, and Daniel M. Croymans (2022) "When peer comparison information harms physician well-being." *Proceedings of the National Academy of Sciences*, 119 (29): e2121730119. Author contributions are: J.S.R., J.C.Z., J.G., H.D., N.M.P., R.K.L., G.G., C.R.F., M.H., and D.M.C. designed research; J.S.R., J.C.Z., J.G., H.D., N.M.P., and D.M.C. performed research; J.S.R. and S.V., analyzed data; J.S.R. and J.C.Z. wrote the paper; and J.S.R., J.C.Z., J.G., H.D., N.M.P., S.V., R.K.L., G.G., C.R.F., M.H., and D.M.C. revised the paper.

Chapter 3 is a paper under second round revision at the *Journal of Marketing Research*. The field experiment was designed and implemented joint with Hengchen Dai, Jana Gallus, Anita McClough, Steve Eitniear, Michelle Slick, and Charlotte Blank. The remaining empirical work, theorizing, and writing was joint with Hengchen Dai and Jana Gallus.

Chapter 4 is a working paper. The empirical work, theorizing, and writing was joint with Jonathan Bogard, Eugene Caruso, and Hal Hershfield.

The figure titles, table titles, and some section headers have been modified for the dissertation.

CURRICULUM VITAE

JOSEPH S. REIFF

EDUCATION

University of California, Los Angeles Los Angeles, CA
Ph.D., Management, expected 2023

Tufts University Medford, MA
B.A., *Summa Cum Laude*, Economics and Psychology, received May 2015

RESEARCH INTERESTS

Social cognition, financial decision making, health behavior, moral judgment, behavioral policy

PUBLICATIONS (see selected abstracts in Appendix)

In all sections, * denotes equal authorship.

Reiff, J. S., Dai, H., Beshears, J., Milkman, K. L., & Benartzi, S. (2022). Save more today or tomorrow: The role of urgency in pre-commitment design. Forthcoming, *Journal of Marketing Research.*

***Reiff, J. S.**, *Zhang, J., Gallus, J., Dai, H., Pedley, N., Vangala, S., Leuchter, R., Goshgarian, G., Fox, C. R., Han, M., & Croymans, D. (2022). When peer comparison information harms physician well-being. *Proceedings of the National Academy of Sciences of the United States of America, 119*(29).

Gallus, J., **Reiff, J. S.**, Kamenica, E., & Fiske, A. P. (2021). Relational incentives theory. *Psychological Review*, *129*(3), 586-602.

Reiff, J. S., Hershfield, H. E., & Quoidbach, J. (2020). Identity over time: Perceived similarity between selves predicts well-being 10 years later. *Social Psychological and Personality Science*, *11*(2), 160-167.

Remedios, J. D., **Reiff, J. S.**, & Hinzman, L. (2020). An identity-threat perspective on discrimination attributions by women of color. *Social Psychological and Personality Science*, *11*(7), 889-898.

Hershfield, H. E., John, E. M., & **Reiff, J. S.** (2018). Using vividness interventions to improve financial decision making. *Policy Insights from the Behavioral and Brain Sciences*, *5*(2), 209-215.

Pepall, L., & **Reiff, J. S.** (2017). Targeted advertising and cumulative exposure effects: The impact of banning advertising to children in Quebec. *Review of Industrial Organization*, *51*(3), 235-256.

Pepall, L., & **Reiff, J. S.** (2016). The Veblen effect, targeted advertising, and consumer welfare. *Economics Letters*, *145*, 218-220.

PAPERS UNDER REVIEW

Reiff, J. S., Dai, H., Gallus, J., McClough, A., Eitniear, S., Slick, M., & Blank, C. (2022). When impact appeals backfire: Evidence from a multinational field experiment and the lab. Invited revision at the *Journal of Marketing Research.*

SELECTED RESEARCH IN PROGRESS

*Bogard, J. E., ***Reiff, J. S.**, Caruso, E. M. & Hershfield H. E. Making sense of dominated options: Implications of dominated options for trust and choice.

Reiff, J. S., Shu, S., Hershfield, H., & Benartzi, S. Using when- (vs. whether-) framing to motivate saving behavior.

Reiff, J. S., Rudkin, A. F., Sundby, S., & Hershfield, H. E. Perspective-taking in capital punishment decisions.

ACADEMIC AWARDS & HONORS

Behavioral Science and Policy Association Best Paper, 2023
Dissertation Year Fellowship, The University of California, 2022-2023
Research Grant, Morrison Center for Marketing and Data Analytics Grant, 2019-2022 (4 grants)
Graduate Research Fellowship, National Science Foundation, 2017-2022
Class of 1942 Scholarship Prize, Tufts University, 2015
Phi Beta Kappa, Tufts University, 2015

INVITED TALKS

Office of Evaluation Sciences, General Services Administration	March 2023
University of Maryland, Robert H. Smith College of Business	February 2023
University of Arizona, Eller College of Management	November 2022
University of Chicago, Booth School of Business	November 2022
University of California, Berkeley, Haas School of Business	October 2022
University of Michigan, Ross School of Business	October 2022
University of Illinois Urbana-Champaign, Gies College of Business	October 2022
Duke University, Fuqua School of Business	October 2022
University of British Columbia, Sauder School of Business	September 2022

TEACHING EXPERIENCE

UCLA Anderson School of Management, Los Angeles, CA

- Teaching Assistant for Marketing Management, Spring 2021
- Teaching Assistant for Choice Architecture in Practice, Winter 2023
- Teaching Assistant for Behavioral Economics, Spring 2023

WORK EXPERIENCE

ThrivePass, *Research Consultant*	2022-present
Federal Reserve Bank of Chicago, *Associate Economist*	2015-2017

INTRODUCTION

Imagine you visit your doctor for an annual checkup, and they ask you to consider getting a painful preventive screening exam. You might be wondering: how urgent is this screening exam? The doctor offers you an appointment for the exam either later today *or* in two months. You infer from the doctor's "now or later" offer that the exam cannot be that urgent; Afterall, they seem to be equally endorsing the "do it later" and "do it now" options. What if, instead, the doctor first offered a same-day appointment. Then only after you declined that offer, they offered the appointment in two months. You might then infer that the doctor thinks this exam is quite urgent. The example illustrates how the arrangement of the doctors' offer implicitly communicates information about the urgency of their recommendation. In this interpersonal context, the decision maker could of course just ask the doctor about the urgency of the exam. However, in policy contexts, where a large organization is implementing a policy, direct communication with the policymaker is less feasible. Instead, the key premise of my dissertation is that *implicit communication*—through details in how a policy is designed and implemented—is a key source of information for decision makers and an important determinant of policy effectiveness.

This premise is supported by research from social psychology (on "intervention construal"; Paluck and Shafir 2017), marketing (on "marketplace metacognition"; Wright 2002), and management (on "sensemaking"; Weick, Sutcliffe, and Obstfeld 2005; Krijnen, Tannenbaum, and Fox 2017). My dissertation seeks to broaden the scope and impact of this work by identifying a series of consequential social inferences that individuals draw from the use of behavioral policies. By combining large-scale natural field experiments with survey data, qualitative responses, and incentive compatible lab experiments, my dissertation attempts to document how these inferences can help explain the effects of popular behavioral policies.

My dissertation also contributes to a growing body of research that studies psychological phenomena in the field. Behavioral policies often have unexpected and inconsistent results across field experiments (List 2022). To explain this variability, recent research has largely focused on differences in sampling and measurement (Saccardo et al. 2023). My dissertation explores a complementary explanation. When policymakers attempt to leverage a single positive mechanism to motivate behavior change, seemingly innocuous details in the design of the intervention may inadvertently introduce a second mechanism—a social inference about the policymaker. Differences in the design and implementation of the same intervention across contexts can thus result in differences in the direction and magnitude of the social inference mechanism. Accordingly, the relative magnitude of the two mechanisms in a given context will determine the effect of the policy on behavior (see also Goswami and Urminksy 2022). Identifying hidden mechanisms can further help researchers redesign interventions to “turn off” unintended second mechanisms that may be inhibiting the effectiveness of interventions. In doing this, researchers can more accurately assess the extent to which the remaining active mechanism drives human behavior in the field.

Each of the four chapters in my dissertation examines an inference that people make about a behavioral policy: inferences about the urgency of policymakers’ recommendations who offer pre-commitment (Chapter 1), inferences about support from policymakers who use peer comparison information (Chapter 2), inferences about inauthenticity of policymakers’ messages that emphasize consumer impact (Chapter 3), and inferences about the trustworthiness of policymakers who include dominated options in choice sets (Chapter 4). The theories I developed around these inferences can help:

(1) explain when policies—even those that were thought to be "silver bullets"—can fail and even backfire (Chapters 1-4)

(2) inform the redesign of psychological interventions so that they have their intended positive effects (Chapters 1 and 3)

(3) identify moderators that help explain the divergent effects of interventions across contexts (Chapter 3)

(4) implicate new outcomes, unexplored in previous research, that are consequential for individuals and organizations (Chapters 2 and 4)

I conclude by discussing limitations and future directions. Together my dissertation makes theoretical contributions to work on social inference making, behavioral policy, and behavior change; meta-scientific contributions to the science of studying psychological phenomena in the field; and practical contributions for policymakers hoping to design and implement behavioral policies that achieve their intended effects.

REFERENCES

Goswami, Indranil and Oleg Urminsky (2021) "Why Many Behavioral Interventions Have Unpredictable Effects in the Wild: The Conflicting Consequences Problem." Forthcoming in Behavioral Science in the Wild.

List, John A. (2022), *The Voltage Effect*, New York: Currency.

Krijnen, Job M. T., David Tannenbaum, and Craig R. Fox (2017), "Choice architecture 2.0: Behavioral policy as an implicit social interaction," *Behavioral Science & Policy*, 3 (2), i–18.

Paluck, E.L. and E. Shafir (2017), "The Psychology of Construal in the Design of Field Experiments," in *Handbook of Economic Field Experiments*, 245–68.

Saccardo, Silvia, Hengchen Dai, Maria Han, Vangala Sitaram, Hoo Juyea, and Jeffrey Fujimoto (2023), "Field Tested : Assessing the Transferability of Behavioral Interventions."

Weick, Karl E., Kathleen M. Sutcliffe, and David Obstfeld (2005), "Organizing and the process of sensemaking," *Organization Science*, 16 (4), 409–21.

Wright, Peter (2002), "Marketplace Metacognition and Social Intelligence," *Journal of Consumer Research*, 28 (4), 677–82.

CHAPTER 1: SAVE MORE TODAY OR TOMORROW

THE ROLE OF URGENCY IN PRE-COMMITMENT DESIGN

Joseph S. Reiff [a]

Hengchen Dai [a]

John Beshears [b]

Katherine L. Milkman [c]

Shlomo Benartzi [a]

[a] University of California, Los Angeles
[b] Harvard University and NBER
[c] University of Pennsylvania

Acknowledgment: We thank our partners at the universities and at the retirement savings plan administrator for implementing the field experiment. We also thank Alessandro Previtero for his helpful comments and Hae Nim Lee, Predrag Pandiloski, and Christine Nguyen for excellent research assistance.

Financial Disclosure: This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1650604. We acknowledge financial support from the National Institutes of Health (grants P01AG005842 and P30AG034532), the Social Security Administration (grant RRC08098400-06-00 to the National Bureau of Economic Research as part of the SSA Retirement Research Consortium), the Pension Research Council/Boettner Center, the TIAA Institute, the Wharton School, the Wharton Behavioral Lab, the UCLA Anderson School of Management, and the UCLA Behavioral Lab. The opinions and conclusions expressed are solely those of the authors and do not represent the opinions or policy of NIH, SSA, any agency of the Federal Government, the Pension Research Council/Boettner Center, the TIAA Institute, or the NBER. See the authors' websites for lists of outside activities.

Abstract: To encourage farsighted behaviors, past research suggests that marketers may be wise to invite consumers to pre-commit to adopt them “later”. However, the authors propose that people will draw different inferences from different types of pre-commitment offers, and that these inferences can help explain when pre-commitment is effective at increasing adoption of farsighted behaviors and when it is not. Specifically, the authors theorize that simultaneously offering consumers the opportunity to adopt a farsighted behavior now or later (i.e., offering “simultaneous pre-commitment”) may signal that the behavior is not urgently recommended; however, offering consumers the opportunity to adopt that behavior immediately and then, only if they decline, inviting them to adopt it later (i.e., offering “sequential pre-commitment”) may signal just the opposite. In a multi-site field experiment (N=5,196), the authors find that simultaneously giving consumers the chance to increase their savings now or later *reduced* retirement savings. Two pre-registered lab studies (N=5,080) show that simultaneous pre-commitment leads people to infer that taking action is not urgently recommended, and such inferences predict less adoption of recommended behaviors. Importantly, offering sequential pre-commitment increases inferred urgency, predicting greater adoption. Together, this research advances knowledge about the limits and potential of pre-commitment.

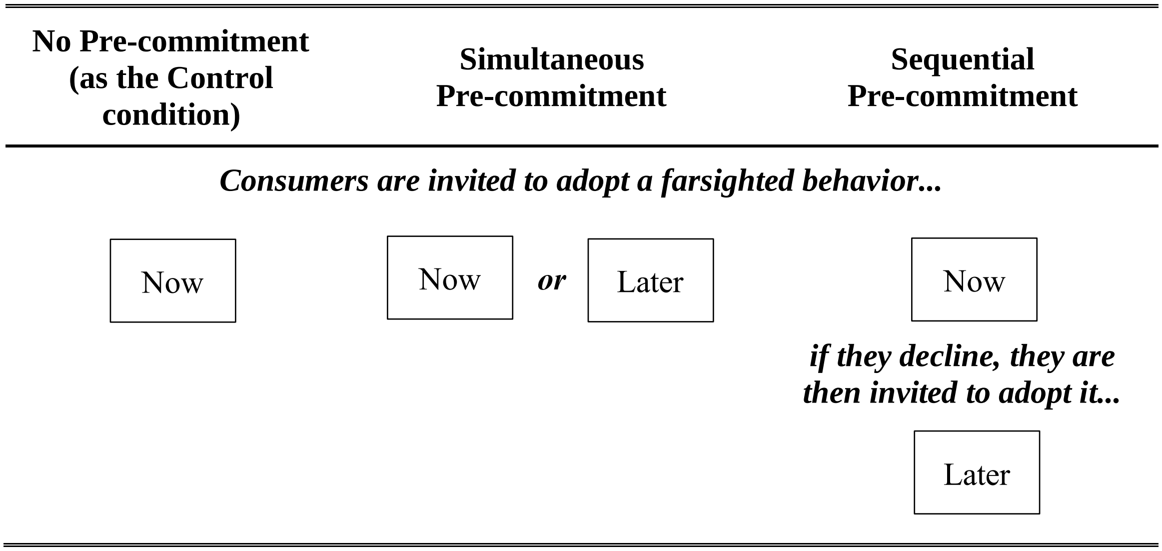
Keywords: pre-commitment, inference making, farsighted decisions, choice architecture, field experiment

INTRODUCTION AND THEORY

Consumers often face decisions about whether and when to engage in behaviors that have immediate costs and long-term benefits (e.g., saving for retirement, updating malware protection software, undergoing preventative health screenings, and receiving a vaccine). When invited to adopt such farsighted activities immediately, people often decline because the immediate costs loom large relative to the distant benefits (Frederick, Loewenstein, and O'Donoghue 2002). Marketers and policy makers seeking to encourage farsighted choices frequently offer "pre-commitment" as a solution: By inviting people to pre-commit to adopt the behavior in the future, it makes the costs feel less aversive, which extant theory and evidence suggest should increase adoption (Milkman, Rogers, and Bazerman 2009; Milkman, Rogers, and Bazerman 2010; Read and Van Leeuwen 1998; VanEpps, Downs, and Loewenstein 2016). For example, companies like Apple and Zoom invite customers who are due for a large software update to install their updates later; Wikipedia offers the option to pre-commit to donate later; and Stickk.com (a popular goal-setting website) allows users to begin goal pursuit on a future date. Across hundreds of employers, pre-commitment to saving is offered through the Save More Tomorrow program, which allows employees to commit now to start saving for retirement in the future (Benartzi and Thaler 2013; Thaler and Benartzi 2004). While previous research suggests that pre-commitment should unambiguously increase overall take-up of farsighted behaviors, we propose that the effects of pre-commitment on the adoption of farsighted behaviors may be more nuanced, and we present evidence consistent with our theorizing.

The current research begins with the premise that there are different ways that a marketer can offer consumers the opportunity to pre-commit to a farsighted behavior. Previous research has focused on single-option choice sets, asking how offering the option to adopt a farsighted behavior at a delay (vs. now) impacts take-up (e.g., Rogers and Bazerman 2008). However, when pre-commitment is offered in practice, marketers commonly offer each consumer both the option to adopt the behavior "now" and the option to adopt the behavior "later". Marketers do this presumably because it allows firms to satisfy consumers' heterogeneous preferences (i.e., attracting consumers who prefer delay *in addition to* those who prefer to start immediately), while expediting the adoption of offered behaviors. In this paper, we propose the first theory to examine two common strategies that marketers use to offer both the pre-commitment option and the option to adopt the behavior immediately. We term these two strategies "simultaneous pre-commitment" and "sequential pre-commitment" (illustrated in Figure 1-1). We further identify a novel mechanism that helps explain the differential effects of simultaneous and sequential pre-commitment on farsighted decisions.

Figure 1-1. Summary of the Three Choice Sets Studied in this Paper.



When offering simultaneous pre-commitment, marketers present the option to adopt a farsighted behavior now and the pre-commitment option *side-by-side*. When Zoom offers their software update, for instance, they offer the option to update it now or at a future date, and these options are offered simultaneously, side-by-side. Similarly, when Wikipedia solicits donations, they simultaneously offer consumers the options to donate “now” or “later”. Indeed, in surveys we conducted, the majority of industry professionals (73.8% of N=229), as well as marketing professors (62.4% of N=85), predicted that offering simultaneous pre-commitment to encourage a policy-relevant behavior—enrollment in a retirement plan—would lead individuals to save more compared to only offering the option to enroll now (see Web Appendix A for more details). However, in the current research, we find that simultaneous pre-commitment does not necessarily increase adoption of farsighted behaviors and can even backfire, leading people to delay important behaviors they would otherwise have engaged in immediately. Such delays can be costly, as waiting longer to begin saving reduces accumulated wealth, waiting longer to update software increases the likelihood of malware attacks, waiting longer for a health screening reduces the likelihood of detecting a disease early enough to cure it, and waiting longer for a vaccine reduces the likelihood of having protection at the time of disease exposure. We theorize and show that simultaneous pre-commitment signals that the marketer who designed the adoption opportunity does not view the offered behavior as very urgent. It is as if the marketer is saying either now or later will suffice.

When offering sequential pre-commitment, marketers first give consumers the option to adopt the farsighted behavior immediately, and then, only if the initial offer is declined, do they offer the option to adopt the behavior later. For example, in one of the original implementations of the Save More Tomorrow retirement savings program, employees were only offered the option to pre-commit to save in the future if they had already rejected an offer to start saving immediately (Thaler and Benartzi 2004). In contrast to simultaneous pre-commitment, we theorize that sequential pre-commitment heightens urgency: By offering immediate adoption before the option to delay the action, the marketer is signaling that they prefer the action be taken sooner rather than later. We further show that this inference about urgency can help explain why sequential pre-commitment effectively increases adoption of farsighted behaviors.

In the remainder of this paper, we first develop our theory, position it in the literature, and motivate our hypotheses. We then present evidence supporting our theory from one large-scale field experiment, one vignette-based laboratory experiment, and one incentive compatible laboratory experiment. We end with a discussion of the implications of our research.

THEORY

Previous Theories of Pre-commitment

In this paper, when we state that people are offered "pre-commitment", we mean that they are offered an option to commit now to adopt a behavior at a future point in time. We consider pre-commitment to be a specific type of "commitment device" (Rogers, Milkman and Volpp 2014) because, broadly speaking, when people are offered a commitment device, they are offered the option to commit to restrict a future choice set (Ashraf, Karlan, and Yin 2006; Schwartz et al. 2014). Our research specifically examines pre-commitment.

Multiple streams of research have found that offering consumers a choice to pre-commit to a farsighted behavior increases take-up. For instance, when making choices for the future (vs. for now), people are more likely to select healthy foods (Milkman, Rogers, and Bazerman 2010; Read and Van Leeuwen 1998; VanEpps, Downs, and Loewenstein 2016), rent educational films (Milkman, Rogers, and Bazerman 2009; Read, Loewenstein, and Kalyanaraman 1999), and support policies that bolster environmental sustainability (Rogers and Bazerman 2008).

The most common explanation for this pattern is that many people tend to exhibit present bias (O'Donoghue and Rabin 1999). Present-biased consumers struggle to make farsighted decisions because they over-weight the immediate costs associated with such behaviors (e.g., the taste disadvantages of healthy food, the discomfort of a vaccine or colonoscopy) while steeply discounting the future benefits (e.g., longevity). Previous research has argued that pre-commitment offers should be attractive to present-biased consumers; if the farsighted behavior (e.g., eating a healthy diet, receiving a shot) starts or occurs in the future, the disutility of incurring the costs will be heavily discounted—making the behavior seem less aversive (Milkman, Rogers, and Bazerman 2008). Accordingly, as long as a decent share of consumers are present-biased, offering them a chance to adopt a farsighted behavior in the future (i.e., to pre-commit) should, on average, increase overall take-up of the farsighted behavior.

In addition to present bias, theories about resource slack and pain of payment lead to a similar prediction that pre-commitment should increase adoption of farsighted behaviors. That is, people expect to have more discretionary resources (i.e., "resource slack") in the future than they do in the present (Zauberman and Lynch 2005), and people find spending resources less painful when they have more resources (Morewedge, Holtzman, and Epley 2007). As a result, people should anticipate that, compared to adopting a farsighted behavior now, pre-committing to adopt it in the future will be less painful and thus more attractive.

When a consumer is offered both the option to adopt a behavior "now" and the option to adopt the behavior "later", we still assume that some combination of the aforementioned mechanisms identified in previous work should make the pre-commitment offer attractive to some extent. However, our focus is on an additional mechanism that has been neglected by extant theory about pre-commitment and may counterbalance these previously studied benefits.

Precommitment and Inferred Urgency

We argue that to understand when pre-commitment fails it is necessary to consider people's inferences about the marketer offering pre-commitment. Generally speaking, past research has shown that consumers make inferences about marketers' motives and recommendations (Kardes, Posavac, and Cronley 2004), particularly based on the options they offer and the way those options are arranged (Benartzi 2001; Lieberman, Duke, and Amir 2019; Krijnen, Tannenbaum, and Fox 2017). For example, people assume the option that marketers set as a "default" is what they recommend (Brown and Krishna 2004; McKenzie, Liersch, and Finkelstein 2006).

Extending this work, we propose that people make inferences about the *urgency* of marketers' implicit recommendations. In previous marketing research, urgency has typically been defined as an objective characteristic of tasks (often referred to as "task urgency"; Zhu, Yang, and Hsee 2018). Tasks with upcoming deadlines, for instance, have greater task urgency than those with more distant deadlines. We argue that urgency can also describe a person's subjective judgment that it is better to take action sooner rather than later. Importantly, consumers may make sense of how urgent a marketer thinks it is for them to take action based on the marketer's implicit and explicit recommendations. For instance, if a financial advisor recommends that her clients start saving *immediately*, the clients may perceive that the financial advisor considers saving to be an urgent priority. In the current research, we examine how consumers make inferences about the urgency of marketers' recommendations from the presentation of choices marketers offer. We define "inferred urgency" as the inference by a consumer that a marketer recommends adoption of a behavior sooner rather than later. Below, we first theorize about how the design of a pre-commitment offer affects consumers' inferred urgency and then hypothesize about the consequences of this for consumer choice.

We specifically propose that consumers make inferences about the urgency of a marketer's implicit recommendation from the order in which a pre-commitment option is presented. Previous research has shown that consumers hold "position-based beliefs" (Valenzuela and Raghubir 2009). Consumers assume that options presented at eye-level, for instance, are placed there by the retailer because they are more popular than the options on the bottom of the shelf (Valenzuela and Raghubir 2009). Building on this work, we argue that consumers do not just hold position-based beliefs about *where* options are placed in a display but also make inferences based on *when* marketers present options in a sequence. We specifically propose that consumers will view the temporal ordering of options as an intentional decision by the marketer that signals how strongly the marketer recommends certain options relative to salient alternatives.

We argue that the two common ways marketers design pre-commitment offers—simultaneously and sequentially—send contrasting signals about the urgency of a marketer's recommendation. When a marketer offers consumers simultaneous pre-commitment, the options to adopt the behavior "now" and "later" are presented side-by-side in the same menu. In presenting these options side-by-side, the marketer does not signal a clear preference between the options. Without additional information, consumers may naturally infer that a marketer endorses both options equally (Fox, Ratner, and Lieb 2005; Tannenbaum, Fox, and Goldstein 2013). To consumers, it is as if the marketer is saying, "either doing it now or later will suffice." When a marketer offers consumers sequential pre-commitment, however, they are offering the option to adopt a given behavior "now" first, and only after their offer is rejected do they offer the option to adopt it "later." We propose that this presentation implies an ordinal ranking of the marketer's recommendations. It would be natural for consumers to infer that the marketer is not outright endorsing the “later” option, but rather treating it like a contingency plan to ensure that if consumers do not adopt what the marketer is offering now, they will at least adopt the behavior at some point in the future. To consumers, it is as if the marketer is saying, “you should do this as soon as you can!”

We formally hypothesize that:

> **Hypothesis 1:** Compared to not offering a pre-commitment option, offering simultaneous pre-commitment will decrease consumers’ inferences about the urgency with which a behavior’s adoption is recommended, whereas sequential pre-commitment will increase the inferred urgency.

Inferred Urgency and Farsighted Decisions

When people infer a behavior is urgently recommended, this should subsequently influence their decisions. Previous work has shown that consumers’ choices are often influenced by their inferences about marketers’ recommendations (Smith, Goldstein, and Johnson 2013). For instance, labeling health care plans “gold,” “silver,” and “bronze” conveys to consumers what marketers consider to be the best, middle, and worst plans, which alters consumers’ insurance choices, even when the labels are assigned arbitrarily (Ubel, Comerford, and Johnson 2015). And, when people infer that a default option is recommended by the marketer, they are typically more likely to choose that option (McKenzie, Liersch, and Finkelstein 2006). Other work has shown that when marketers communicate the urgency of a task by highlighting an upcoming deadline, people are more motivated to do it (d’Adda, Galliera, and Tavoni 2020; Zhu, Yang, and Hsee 2018).

Bridging this work, we argue that when consumers infer a behavior is urgently recommended, they should be more likely to adopt the behavior. In the context of pre-commitment offers, our research focuses on inferred urgency's influence on two choice outcomes: immediate adoption and overall adoption of a farsighted behavior. Both of these outcomes have important consequences for consumer well-being, and the two outcomes together present a comprehensive evaluation of how pre-commitment design affects engagement in farsighted behaviors.

Immediate adoption. Immediate adoption refers to whether consumers commit to adopt a behavior immediately (i.e., choosing the "do it now" option). When studying farsighted behaviors, immediate adoption is particularly important to examine because the benefits of these behaviors typically accumulate over time. For instance, saving earlier (vs. later) in life results in greater accumulated savings; updating software sooner (vs. later) increases likelihood of stopping a malware attack; getting screening exams sooner (vs. later) increases likelihood of catching a disease early enough to cure it, and receiving a vaccine sooner (vs. later) increases the likelihood of being protected at the time of disease exposure.

When consumers infer a behavior is urgently recommended by a marketer, they believe that the marketer recommends they adopt the behavior sooner rather than later, which should in turn increase *immediate adoption* of the behavior. Given that simultaneous pre-commitment signals lower urgency (compared to no pre-commitment or sequential pre-commitment) per our earlier theorizing, and given that this lack of urgency should reduce immediate adoption, we expect simultaneous pre-commitment to decrease immediate adoption of farsighted behaviors.[1] More formally, we hypothesize that:

---

[1] Sequential pre-commitment and no pre-commitment (the control condition) should have the same level of immediate adoption by design, as these two conditions are identical until after consumers make decisions about immediate adoption (see Figure 1-1). Thus, we do not make predictions about the effect of sequential pre-commitment (vs. no pre-commitment) on immediate adoption of farsighted behaviors. This also presents a minor exception to Hypothesis 4 (introduced later in the Theory section): we do not predict that inferred urgency mediates the effects of sequential pre-commitment (vs. no pre-commitment) on immediate adoption of farsighted behaviors.

Hypothesis 2: Offering simultaneous pre-commitment will decrease the immediate adoption of farsighted behaviors compared to offering sequential pre-commitment or making no pre-commitment offer.

Overall adoption. Overall adoption refers to whether consumers commit to adopt a behavior at any point in time (i.e., choosing either the "do it now" option or the "do it later" option). When consumers infer that a marketer recommends they promptly adopt a behavior as soon as possible, this should in turn increase *overall adoption* (i.e., prompting them to commit to do it, either immediately or at a future time).

The predicted impact of simultaneous pre-commitment (relative to no pre-commitment) on overall adoption is unclear due to competing mechanisms. On one hand, simultaneous pre-commitment may capitalize on people's preference for delaying the adoption of farsighted behaviors (as found in previous research), which should increase overall adoption by getting people who otherwise would not sign up to choose the pre-commitment option. On the other hand, it reduces inferred urgency, which should curb adoption of the farsighted behavior. Therefore, we do not make predictions about the effect of simultaneous pre-commitment (vs. no pre-commitment) on overall adoption of farsighted behaviors because the effect will depend on the relative strength of these opposing mechanisms.

However, sequential pre-commitment should unambiguously increase overall adoption of farsighted behaviors relative to not offering a pre-commitment option, because it both signals greater urgency and capitalizes on people's preference to pursue farsighted behaviors at a time delay. Further, compared to offering simultaneous pre-commitment, offering consumers sequential pre-commitment should also clearly boost overall adoption: both types of pre-commitment leverage people's preference for delaying the costs associated with farsighted activities, but we expect sequential pre-commitment to signal a greater sense of urgency than simultaneous pre-commitment. We therefore hypothesize that:

> **Hypothesis 3:** Sequential pre-commitment will increase overall adoption of farsighted behaviors, compared to making no pre-commitment offer or offering simultaneous pre-commitment.

Mediation via Inferred Urgency. Given our predictions that simultaneous and sequential pre-commitment impact consumers' inferences about how urgently action is recommended and that heightened inferred urgency spurs immediate and overall adoption of farsighted behaviors, we formally hypothesize:

> **Hypothesis 4**: Inferred urgency will mediate the effects of simultaneous and sequential pre-commitment (vs. not offering a pre-commitment option) on both the immediate and overall adoption of farsighted behaviors.

The four hypotheses are summarized in Table 1-1.

Table 1-1. Summary of the Theory's Key Predictions.

Measure	**Effects of Simultaneous Pre-commitment**	**Effects of Sequential Pre-commitment**
	(vs. No Pre-commitment)	
Inferred Urgency	Decrease (H1)	Increase (H1)
Immediate Adoption of Farsighted Behaviors	Decrease (H2) Because of Decreased Urgency (H4)	No Effect Because Sequential Pre-commitment and No Pre-commitment Are Identical up to the Point of the Immediate Adoption Decision[a]
Overall Adoption of Farsighted Behaviors	Ambiguous Effect Because Decreased Urgency (H4) May be Offset by Counteracting Mechanisms[b]	Increase (H3) Because of Increased Urgency (H4)

[a] For more information see footnote 1. [b] For more information see the theory section on overall adoption.

Theoretical Implications

Overall, the current paper aims to make three main contributions. First, we contribute to research on pre-commitment and farsighted decision making by drawing a theoretical and practical distinction between simultaneous and sequential pre-commitment. Although commonly used in practice and thus worthy of systematic investigation, these types of pre-commitment have not previously been distinguished from one another and rigorously studied.

Second, we contribute to research on inference making (Kardes, Posavac, and Cronley 2004) and information leakage (McKenzie, Liersch, and Finkelstein 2006) by uncovering a novel, consequential inference people draw from choice sets: the inferred urgency of the marketer's implicit recommendation. We theorize about why people form inferences about the urgency of recommendations and why inferred urgency can spur people to take prompt action. Further, we argue and show that inferred urgency can help explain when offering pre-commitment increases adoption of farsighted behaviors and when it does not.

Third, we present a large, real-world experimental test of pre-commitment, arguably one of the most commonly used "nudge" interventions. Our research suggests that a seemingly small difference in the way a popular idea is implemented (e.g., the simultaneous vs. sequential presentation of a pre-commitment option) can change its effects. We discuss generalizable lessons for scaling promising marketing strategies in the field.

The remainder of this paper is organized as follows. We begin our investigation with a field experiment studying simultaneous pre-commitment, and then we present two additional well-powered, pre-registered laboratory studies testing all four of our hypotheses. The results of each study are summarized in Table 1-2. Our pre-registrations, materials, non-proprietary data, and code are available at: https://researchbox.org/434.

Table 1-2. Summary of Key Results.
We report raw means and standard deviations in parentheses. Statistically significant differences between conditions are indicated by stars (* $p \leq .10$, ** $p \leq .05$, *** $p \leq .01$) and were estimated using the primary analytical model specified in each study. Checkmarks in the far-right column indicate that inferred urgency mediates all reported differences in adoption between conditions.

	Sample & Measures	**Details**	**No Pre-commitment**	**Simultaneous Pre-commitment**	**Sequential Pre-commitment**	**Mediation via Inferred Urgency**
Study 1	Sample	N = 5,196 university employees	n = 2,600	n = 2,596		
			***			
	Immediate adoption	Whether employees enrolled in a savings plan immediately (=1), or not (=0)	.075 (.26)	.059 (.23)		
			n.s.			
	Overall adoption	Whether employees enrolled in a savings plan immediately or at a delay (=1), or not at all (=0)	.123 (.33)	.116 (.32)		
Study 2	Sample	N = 2,682 participants from Prolific	n = 901	n = 895	n = 886	
				***		
			***	***		
	Inferred urgency	Inference about the urgency of the employer's recommendation (1-7 response scale)	4.88 (1.52)	4.55 (1.63)	5.22 (1.30)	
			***	***		
	Immediate adoption	The number of employee benefits participants enrolled in immediately (ranging from 0-3)	1.86 (.78)	1.49 (.89)	1.93 (.85)	✓
				***		
			n.s.	***		
	Overall adoption	The number of employee benefits participants enrolled in immediately or at a delay (ranging from 0-3)	1.86 (.78)	1.84 (.81)	2.14 (.82)	✓
Study 3	Sample	N = 2,398 participants from MTurk	n = 806	n = 794	n = 798	
				***		
			***	***		
	Inferred urgency	Inference about the urgency of the researcher's recommendation (1-7 response scale)	3.87 (1.79)	3.57 (1.89)	4.23 (1.81)	
			*	**		
	Immediate adoption	Whether participants enrolled in a financial assessment immediately (=1), or not (=0)	.320 (.47)	.282 (.45)	.331 (.47)	✓
				***		
			***	***		
	Overall adoption	Whether participants enrolled in a financial assessment immediately or at a delay (=1), or not at all (=0)	.320 (.47)	.480 (.50)	.574 (.49)	✓

STUDY 1: SIMULTANEOUS PRE-COMMITMENT IN THE FIELD

To test our theory's main predictions about the impact of offering simultaneous pre-commitment in the field, we report on the results of two conditions from an experiment involving real savings decisions.[2] A companion paper (Beshears et al. 2021) compares a third condition from this field experiment with one of the conditions examined in our paper to explore a separate research question (see additional details in the following sections and Web Appendix B).

Methods

Four U.S. universities (labeled Universities A, B, C, and D to preserve their anonymity) collaborated with us on our field experiment.[3] Each university began by identifying a retirement savings plan in which they hoped to increase employees' contributions. All universities then identified employees who were not enrolled in this "targeted plan" and therefore had a contribution rate of zero, encouraging them to sign up to save in the targeted plan. One university (University D) also identified employees who were contributing to the targeted plan, but not at the level necessary to take full advantage of their employer's matching contributions; these employees were encouraged to save more, rather than to start saving. Table 1-3 presents more information about the targeted plans offered by the four universities, and Table A1-2 in Web Appendix C details other (non-targeted) savings plans.

---

[2] Note that although the field experiment fits as a test of our theory's predictions, it was conducted before the development of our theory. For full transparency, we were originally hoping, based on prior research, that offering a pre-commitment option would increase retirement savings. We were surprised by the negative impact of our pre-commitment design on savings, which prompted us to develop our theory and pre-register Studies 2-3 to deductively test our theory.

[3] The experiment originally included a fifth university. However, this university requested that employees elect dollar contribution amounts instead of contribution rates. Consequently, this university had different mailing designs from other universities. Further, this university had a very low response rate (only .6% of employees at this university increased their contribution rate by the end of our study period across the Simultaneous Pre-commitment and No Pre-commitment conditions, compared to an average of 13.0% at the other universities). Thus, it was not possible to do a meaningful analysis for this university, and we excluded its data from our analysis.

Table 1-3. Descriptions of Targeted Plans.
The table is identical to the one presented in Beshears et al. (2021) because it describes the same retirement plans in the same universities.

University	Eligibility	Employer Contributions
A	All employees on the University's payroll with FICA deductions	None
B	All employees whose annual contribution limit to the targeted plan is at least \$200	None
C	All paid employees OR students with a stipend	None
D	*Eligibility for Employee Contributions*	*Automatic Employer Contribution Rates (Regardless of Whether the Employee Contributes)*
	i) Regular full-time staff (with monthly or weekly pay cycles) OR ii) Full-time faculty and academic support staff in a benefits-eligible title OR iii) Limited-service staff scheduled to work at least 35 h per week for a minimum of 9 months per year (with monthly or weekly pay cycles)	i) 1.5% (employee age < 30) ii) 3% (employee age 30–39) iii) 4% (employee age ≥ 40)
	Eligibility for Employer Contributions All employees who are eligible for employee contributions (described above), are age 21 or older, and have at least one year of prior service	*Matched Employer Contributions* Dollar-for-dollar match on employee contributions up to 5% of employee's salary

One retirement plan record keeper shared by all four universities sent out mailings in early October of 2013 to university employees' homes. The mailings provided employees with an opportunity to increase their savings contributions by filling out and mailing back a simple form on a pre-stamped, pre-addressed postcard. If an employee checked a box indicating they wanted to save and then signed and returned the postcard, that employee would be enrolled in the plan at a preselected contribution rate with their contributions allocated to a preselected fund. At all four universities, the preselected fund on the mailing was a lifecycle fund, which provided a diversified portfolio with a mixture of equity, bond, and money market funds tailored to the employee's age. The preselected contribution rate was 3% of the employee's pay for Universities A-C and 5% for University D. If an employee who was already contributing to the targeted plan elected to save more (only relevant to University D), their contributions would increase to the preselected rate with the contributions allocated according to their existing asset allocation.

Employees were randomly assigned to different experimental conditions, which determined the exact mailing they received (see Figure 1-2 and Web Appendix B for additional details on our stratified random assignment process). In this paper, we only analyze employees who were randomly assigned to receive either a No Pre-commitment mailing or a Simultaneous Pre-commitment mailing (N = 5,196; M_{age} = 43.10, SD_{age} = 12.05; 52.3% female). The No Pre-commitment mailing offered employees the opportunity to immediately increase their contribution rate to the targeted plan. The Simultaneous Pre-commitment mailing offered employees the opportunity to increase their contribution rate to the targeted plan either immediately or after a time delay (e.g., "in two months") ranging from two to six months. Both mailings are displayed in Figure 1-3.

Figure 1-2. Randomization.
Employees who received the No Pre-commitment mailings were not included in the Beshears et al. (2021) paper and are only analyzed here. Employees who received the Simultaneous Pre-commitment mailings linked with temporal landmarks were not included in analyses in this paper (this condition is shaded grey). The Simultaneous Pre-Commitment condition linked with temporal landmarks was oversampled because the stratification procedure was designed to allow Beshears et al. (2021) to make inferences about the relative impact of different types of temporal landmarks referenced such as birthdays, the start of spring, new year's, etc. (see details in Web Appendix B).

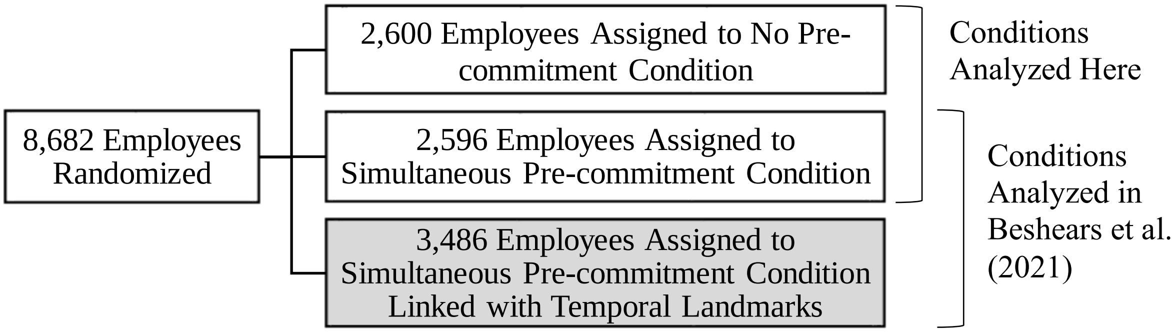
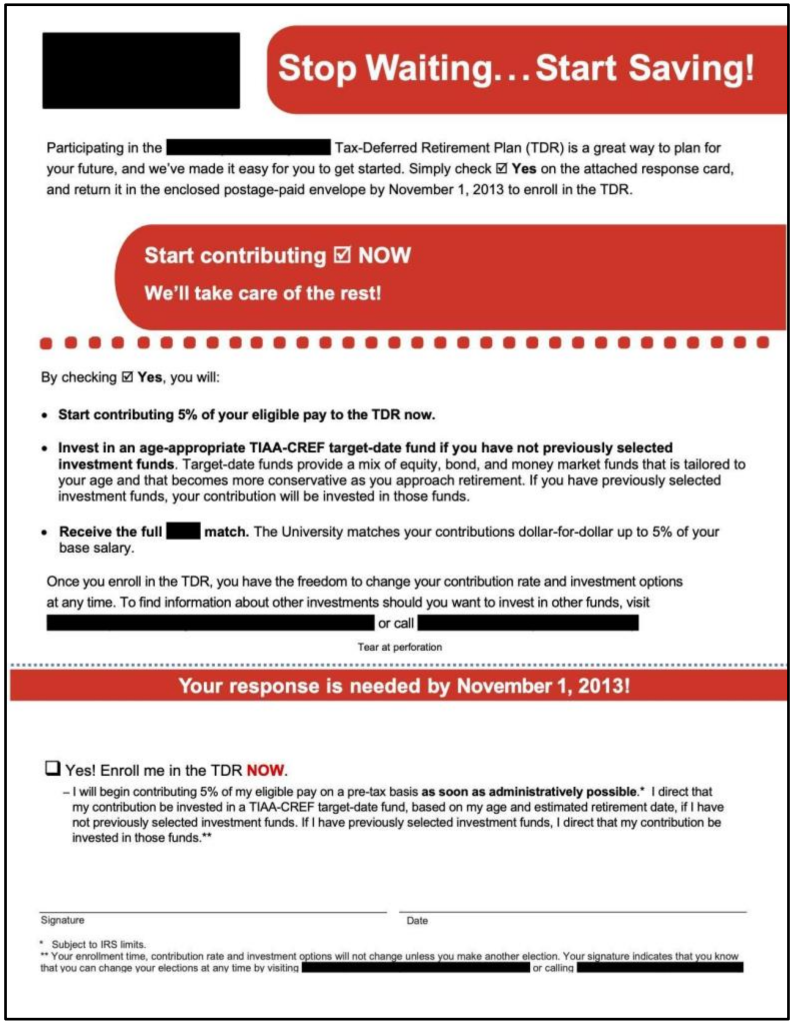


Figure 1-3. Mailings from Field Experiment (Study 1)

No Pre-commitment Mailing



[redacted]

Stop Waiting…Start Saving!

Participating in the [redacted] Tax-Deferred Retirement Plan (TDR) is a great way to plan for your future, and we've made it easy for you to get started. Simply check ☑ **Yes** on the attached response card, and return it in the enclosed postage-paid envelope by November 1, 2013 to enroll in the TDR.

Start contributing ☑ NOW

We'll take care of the rest!

By checking ☑ **Yes**, you will:

- **Start contributing 5% of your eligible pay to the TDR now.**
- **Invest in an age-appropriate TIAA-CREF target-date fund if you have not previously selected investment funds.** Target-date funds provide a mix of equity, bond, and money market funds that is tailored to your age and that becomes more conservative as you approach retirement. If you have previously selected investment funds, your contribution will be invested in those funds.
- **Receive the full [redacted] match.** The University matches your contributions dollar-for-dollar up to 5% of your base salary.

Once you enroll in the TDR, you have the freedom to change your contribution rate and investment options at any time. To find information about other investments should you want to invest in other funds, visit [redacted] or call [redacted]

Tear at perforation

Your response is needed by November 1, 2013!

☐ Yes! Enroll me in the TDR **NOW**.

– I will begin contributing 5% of my eligible pay on a pre-tax basis **as soon as administratively possible.*** I direct that my contribution be invested in a TIAA-CREF target-date fund, based on my age and estimated retirement date, if I have not previously selected investment funds. If I have previously selected investment funds, I direct that my contribution be invested in those funds.**

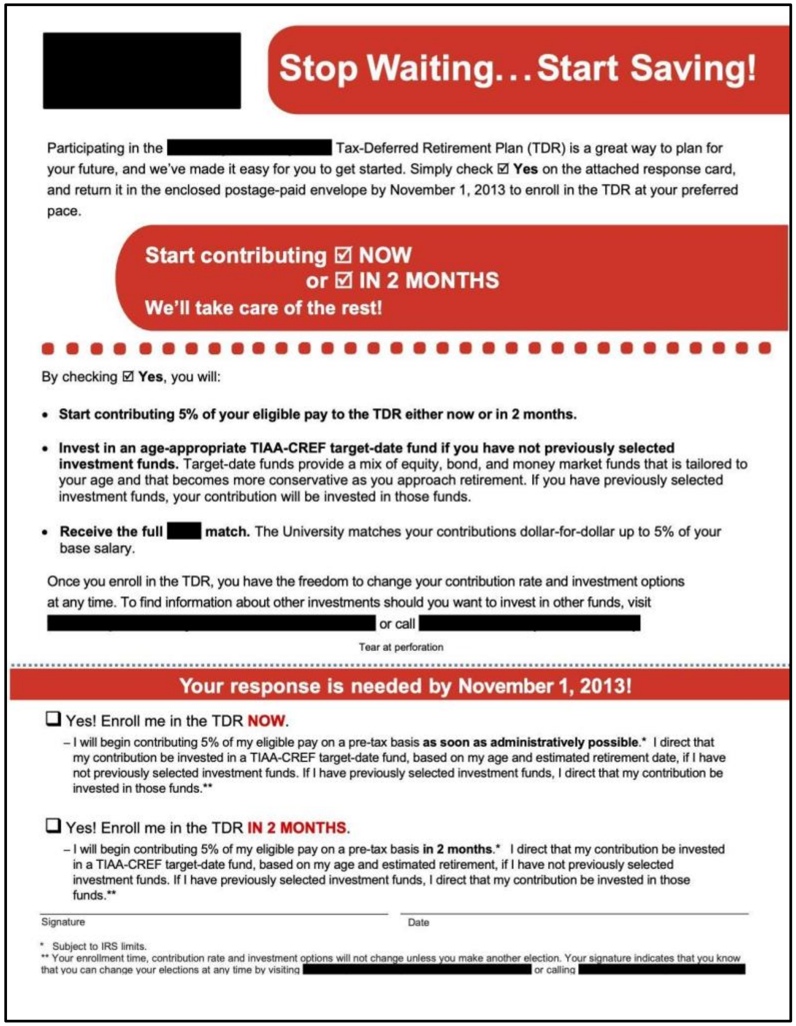
Signature

Date

* Subject to IRS limits.

** Your enrollment time, contribution rate and investment options will not change unless you make another election. Your signature indicates that you know that you can change your elections at any time by visiting [redacted] or calling [redacted]

Simultaneous Pre-commitment Mailing



[redacted]

Stop Waiting…Start Saving!

Participating in the [redacted] Tax-Deferred Retirement Plan (TDR) is a great way to plan for your future, and we've made it easy for you to get started. Simply check ☑ **Yes** on the attached response card, and return it in the enclosed postage-paid envelope by November 1, 2013 to enroll in the TDR at your preferred pace.

Start contributing ☑ NOW
or ☑ IN 2 MONTHS

We'll take care of the rest!

By checking ☑ **Yes**, you will:

- **Start contributing 5% of your eligible pay to the TDR either now or in 2 months.**
- **Invest in an age-appropriate TIAA-CREF target-date fund if you have not previously selected investment funds.** Target-date funds provide a mix of equity, bond, and money market funds that is tailored to your age and that becomes more conservative as you approach retirement. If you have previously selected investment funds, your contribution will be invested in those funds.
- **Receive the full [redacted] match.** The University matches your contributions dollar-for-dollar up to 5% of your base salary.

Once you enroll in the TDR, you have the freedom to change your contribution rate and investment options at any time. To find information about other investments should you want to invest in other funds, visit [redacted] or call [redacted]

Tear at perforation

Your response is needed by November 1, 2013!

☐ Yes! Enroll me in the TDR **NOW**.

– I will begin contributing 5% of my eligible pay on a pre-tax basis **as soon as administratively possible.*** I direct that my contribution be invested in a TIAA-CREF target-date fund, based on my age and estimated retirement date, if I have not previously selected investment funds. If I have previously selected investment funds, I direct that my contribution be invested in those funds.**

☐ Yes! Enroll me in the TDR **IN 2 MONTHS**.

– I will begin contributing 5% of my eligible pay on a pre-tax basis **in 2 months.*** I direct that my contribution be invested in a TIAA-CREF target-date fund, based on my age and estimated retirement, if I have not previously selected investment funds. If I have previously selected investment funds, I direct that my contribution be invested in those funds.**

Signature

Date

* Subject to IRS limits.

** Your enrollment time, contribution rate and investment options will not change unless you make another election. Your signature indicates that you know that you can change your elections at any time by visiting [redacted] or calling [redacted]

The experiment included another group of employees who were randomly assigned to receive a different type of Simultaneous Pre-commitment mailing, which offered them an opportunity to increase their contribution rate to the targeted plan either after a labeled temporal landmark (e.g., their birthday, the first day of spring, Thanksgiving, Valentine's Day) or immediately. A companion paper (Beshears et al. 2021) compares the enrollment decisions of employees who received these distinctive Simultaneous Pre-commitment mailings linked with temporal landmarks and employees who received the standard Simultaneous Pre-Commitment mailings studied here. The objective of Beshears et al. (2021) was to test whether inviting people to boost their contribution rate after a "fresh start" date (e.g., a birthday, the first day of spring; following Dai, Milkman and Riis, 2015), increases savings over and above inviting people to increase contributions at an equidistant future time point (e.g., in 2 months). Beshears et al. (2021) does not report results from the No Pre-commitment condition studied here because the paper solely explores the effect of inviting savings following fresh start dates and not the effects of offering pre-commitment. See Web Appendix B for more information about Beshears et al. (2021).

This experiment's randomization was stratified by university because the universities varied on important features such as the targeted plans' characteristics. Randomization was also stratified by birth month (within each university) because employees' birthdays partially determined which mailing they received; only those whose birthday fell into November 2013-March 2014 had the opportunity to be randomized to receive the option to save more after their birthday.

In all conditions, mailing recipients who wanted to increase their contribution rates had to send back their response card by November 1, 2013. If they chose to save at a higher rate immediately via the mailer, their contribution rate would increase to the preselected rate in November 2013. If they chose to save more at a delay (e.g., in five months), their contribution rate would automatically increase to the preselected rate at the predetermined time (e.g., in March 2014).

Data

Our university partners first pulled a cross-sectional snapshot of information about all plan-eligible employees in August 2013, including their current contributions to the targeted plan, contributions to all other non-targeted savings plans, birth date, hire date, termination date, salary, and position (faculty versus staff). Our conditions are reasonably well balanced across baseline employee characteristics, with the only statistically significant difference being that the mean salary of employees in the No Pre-commitment condition (M = $56,505.19, SD = $35,234.21) is slightly less than that of employees in the Simultaneous Pre-commitment condition (M = $58,505.26, SD = $36,111.88; $p = .043$; See Table A1-1 in Web Appendix B). To ensure that the slight imbalance detected on this dimension is accounted for, our regressions control for baseline employee characteristics, including salary decile.

After the study concluded, our university partners provided information on each employee's contributions to the targeted plan and all other retirement savings plans as well as their pay for each pay cycle from August 2013 through June 2014.[4] We measured the impact of our mailing by observing changes in employees' retirement plan contributions (made by mail, phone or online).

Variables

---

[4] We cannot publish any data from our field experiment due to the nondisclosure data agreement we signed with our field partners. However, if any researcher is interested in replicating our analyses, they should contact us, and we will try to have them added to our nondisclosure data agreement so that individual scholars may be able to work with our field data.

To comprehensively measure the effects of offering simultaneous pre-commitment on savings, we created three outcome variables, which are described below.

Immediate adoption. *Immediate adoption* is a binary variable that takes on a value of one for people who increased their contribution rate to the targeted plan immediately after receiving our mailing and zero for others. We constructed this variable by examining whether an employee's contribution rate in November 2013 (the first month our mailings could have triggered increased contributions) was higher than their rate in September 2013 (the month right before our mailings were sent out).[5]

Overall adoption. *Overall adoption* is a binary variable that takes on a value of one for people who increased their contribution rate to the targeted plan by the end of our study period and zero for others. We constructed this variable by examining whether an employee's contribution rate in June 2014 (the last month in which we received data on employees' contributions and pay) was higher than their contribution rate in September 2013.[6]

Average savings rate. To capture employees' cumulative retirement savings (adjusted for their salary) during our study period, we calculated every employee's *average savings rate* by taking the total number of dollars the employee contributed to the targeted plan from November 2013 through June 2014 and dividing it by the employee's total pay during the same period. This outcome variable ranges from 0 to 1, representing the percentage of an employee's total pay that was contributed to the targeted savings plan during our study period.

---

[5] In the manuscript, we focus on when and how contributions to the targeted plan changed because our mailings encouraged employees to increase savings in the targeted plan. However, as shown in Table A1-4 in Web Appendix E, our effects are robust if we comprehensively examine the impact of our pre-commitment design using employee contributions to *all* savings plans offered by their employer (including the targeted plan).

[6] We also calculated overall adoption using an alternative method that was meant to capture the direct responses to the mailings. Specifically, we only counted someone as enrolling if the first time their contribution rate increased (relative to their rate in September 2013) was either in November 2013 (i.e., the immediate enrollment option) or the specific month when pre-commitment was invited in their mailing. Note that our results do not change substantively when we examine this narrower outcome (see Table 5 and Web Appendix E).

Analysis Strategy

To estimate the causal impact of the Simultaneous Pre-commitment mailing (compared to our No Pre-commitment mailing that only invited people to save now), we relied on the following ordinary least squares (OLS) regression specification:

$$\text{(1)} \qquad outcome_i = \alpha + \beta\ simultaneous\ pre\text{-}commitment_i + \gamma' X_{ij} + \Sigma_j\ (\delta_j\ I[university_i = j] + \zeta_j' X_{ij}\ I[university_i = j]) + \varepsilon_i$$

where i indexes employees and j indexes universities. We estimated this regression once with each of the outcome variables explained above. The coefficient on the indicator for simultaneous pre-commitment is the estimate of the causal impact of the Simultaneous Pre-commitment condition relative to the No Pre-commitment condition. In order to increase statistical power, we estimated a single treatment effect across universities instead of separate treatment effects for each university, but this decision does not invalidate the interpretation of the coefficient as a causal effect, since randomization was stratified by university. X_{ij} is a vector of controls: gender, age decile, tenure decile, salary decile, faculty status, and birth month, where decile breakpoints are calculated separately for each university. $I[university_i = j]$ is an indicator variable that takes a value of one when employee i is associated with university j and a value of zero otherwise.[7] The δ_j and ζ_j coefficients allow the intercept term and the coefficients on the control variables to vary by university, accounting for differences across universities in their average responsiveness to our mailings and differences across universities in the relationship between the control variables and the outcome variable. Given that retirement savings decisions are largely determined by

---

[7] Some employees at University D were already enrolled in the targeted plan before the experiment started, and the experimental mailings encouraged them to further increase their contribution rates. Because these employees are qualitatively different from those who were not yet contributing, in our analyses we treat the two groups of employees as belonging to separate "universities" by including two distinct "university" indicator variables for those two groups. Also, note that in the summation shown in Equation 1, we omit one university indicator variable to avoid collinearity.

socioeconomic circumstances, controlling for employees' demographics, income, and employer characteristics in analyses of interventions designed to increase savings rates can dramatically enhance statistical power and is consistent with standard practice in retirement savings research (for a review, see Choi 2015; see Web Appendix D for more information). We report heteroskedasticity-robust standard errors. For the binary outcomes, we report linear probability regressions here rather than logistic regressions for the ease of interpretation, but the results do not substantively differ when we estimate logistic regressions (see this and all other robustness checks summarized in Table 1-5 and described in detail in Web Appendix E).

Table 1-4. The Effect of Offering Employees Simultaneous Pre-commitment in Study 1. Model 1 reports an ordinary least squares (OLS) regression where the dependent variable is a binary variable reflecting whether an employee immediately increased their contribution rate to the targeted savings plan. Model 2 reports an OLS regression where the dependent variable is a binary variable reflecting whether an employee increased their contribution rate to the targeted savings plan by the end of our study period. Model 3 reports an OLS regression where the dependent variable is an employee's average savings rate in the targeted savings plan during our study period.

	Model 1: Immediate Adoption	Model 2: Overall Adoption	Model 3: Average Savings Rate
Simultaneous Pre-commitment	-.019***	-.009	-.0014**
	(.007)	(.009)	(.001)
University FEs	Yes	Yes	Yes
Controls	Yes	Yes	Yes
University FE x Controls	Yes	Yes	Yes
R-squared	.07	.08	.11
Observations	5,196	5,196	5,196

* $p \leq .10$, ** $p \leq .05$, *** $p \leq .01$

Notes: Standard errors robust to heteroskedasticity are reported in parentheses. Controls include gender, age decile, tenure decile, tenure decile, salary decile, faculty status, and birth month.

Table 1-5. Robustness Checks for the Effect of Offering Simultaneous Pre-commitment in Study 1.
Each row corresponds to a robustness check testing the effects of simultaneous pre-commitment (vs. no pre-commitment). Unless otherwise specified, the models are similar to the primary model specified in the Analysis Strategy section. In the right three columns, we report the coefficients from the regressions in the relevant row, with heteroskasticity robust standard errors in parentheses and significance indicated by * $p \leq .10$, ** $p \leq .05$, *** $p \leq .01$. Full descriptions of each robustness check are reported in Web Appendix E.

Robustness Check		***Results***		
		Immediate Adoption	Overall Adoption	Average Savings Rate
Measuring Decisions with Cards Mailed back to the One University that Tracked Responses. Here we measure immediate and overall adoption by analyzing the choices made on returned response postcards at the one university that shared this information (N=2,029). See Table A1-8.		-.016* (.009)	-.008 (.009)	
Measuring Overall Adoption Differently. To measure overall adoption in a way meant to capture employees' direct responses to our mailers, we only counted someone as enrolling if the first time their contribution rate increased (relative to their rate in September 2013) matched one of the month(s) offered to them in our mailing. See Table A1-9.			-.003 (.007)	
Running Logistic and Fractional Logit Regressions. Here we rely on logistic regression to analyze immediate and overall adoption and fractional logistic regression to analyze average savings rates. The coefficients reported are in terms of log odds ratios (and thus have different interpretations than those from the other robustness checks). See Table A1-10.		-.359*** (.118)	-.099 (.092)	-.163** (.081)
Dropping Those with Missing Data. Here we drop data from employees who have missing data for salary or contributions in all months in the study period (rather than assuming 0's). See Table A1-11.		-.020*** (.007)	-.010 (.009)	-.0015** (.001)
Dropping Those with Missing Data in Key Months. Here we drop data from employees who have missing data for salary or contributions in one of the key months used to calculate immediate and overall adoption (rather than assuming 0's). See Table A1-12.		-.021*** (.008)	-.009 (.010)	
Varying the Cutoff for a Contribution Rate Increase. Here we use different thresholds to measure immediate and overall adoption. Since we constructed the adoption variables by assessing whether increases in contribution rates occurred, here we ensure our results are not spuriously driven by how we rounded contribution rates. See Table A1-13. We counted a contribution rate as having increased if the increase was…	≥ .1% of salary	-.020*** (.007)	-.009 (.009)	
	≥ .01% of salary	-.021*** (.007)	-.008 (.009)	
Including Limited Controls. Here we control for the interaction between university and birth month (due to the stratified random assignment employed) but drop other controls (i.e., gender, age, tenure, salary, and faculty status). See Table A1-14.		-.017** (.007)	-.007 (.009)	-.0012* (.001)

Results

Immediate Adoption. Consistent with Hypothesis 2, employees' probability of immediately increasing their contributions to the targeted plan was 1.9 percentage points lower in the Simultaneous Pre-commitment condition than in the No Pre-commitment condition (p = .006; Table 1-4 Model 1). In terms of its relative effect, this represents a 25.3% decrease, relative to the 7.5 percent of employees who immediately increased their contribution rate to the targeted plan in the No Pre-commitment condition.

Overall Adoption. Importantly, when we look at the full study period through June 2014, there was not a statistically significant difference between conditions in employees' probability of increasing their contributions to the targeted savings plan (p = .301; Table 1-4 Model 2).

Given that simultaneous pre-commitment decreased the number of people who immediately increased their contribution rate to the targeted saving plan but did not significantly affect the total number of savers, what we can conclude is that simultaneous pre-commitment led some people to delay saving who otherwise would have started saving immediately.[8] When people delay saving, they save less overall than they would have if they started saving earlier. We next formally tested whether offering simultaneous pre-commitment ultimately reduced accumulated savings.

Average Savings Rate. Table 1-4 Model 3 indicates that receiving the Simultaneous Pre-commitment mailing (rather than the No Pre-commitment mailing) decreased average saving rates to the targeted plan by .14 percent of pay (p = .046). In relative terms, receiving the Simultaneous Pre-commitment mailing caused a 16.5% decrease in savings in targeted plans

---

[8] We confirmed that, conditional on employees deciding to increase contributions, the Simultaneous Pre-commitment mailing did not affect the magnitude of increases in contribution rates to the targeted plans (p = .311).

during our study period compared with the No Pre-commitment mailing (for which the average savings rate was .85 percent of employee pay).

In additional models reported in Web Appendix E, we examined how the treatment effect on each of the outcome variables varied by university. The effects of simultaneous pre-commitment on both immediate and overall adoption did not significantly vary by university, though the decrease in average savings rates was more prominent in two universities.

Discussion

Contrary to past research demonstrating the benefits of pre-commitment, we find in a large field experiment conducted with four employers that inviting consumers to save more “now or later” (i.e., offering simultaneous pre-commitment) did not increase the share of people contributing to a retirement savings plan. In fact, offering simultaneous pre-commitment (relative to simply inviting consumers to save now) on average decreased overall retirement wealth because some people opted to delay increasing their contribution rates to the savings plan and thus saved over a shorter time horizon than they would have otherwise.

As mentioned earlier, in a companion paper (Beshears et al. 2021), the Simultaneous Pre-commitment condition from this field experiment is used in conjunction with an additional experimental condition not analyzed here to examine the effects of offering pre-commitment shortly after fresh start dates. That paper finds that offering pre-commitment after fresh starts (e.g., birthdays, the first day of spring) increases overall adoption and average savings rates compared to offering pre-commitment after an equidistant future time point that is not associated with such a temporal landmark. The primary contribution of Beshears et al. (2021)—to demonstrate the impact of linking opportunities to save with dates that feel like fresh starts—is theoretically distinct from the key contribution of Study 1 in our paper, which is to show the impact of simultaneous pre-commitment (vs. not offering pre-commitment). See Table A1-15 in Web Appendix F for a summary of how Beshears et al. (2021) and Study 1 in the current paper differ in terms of their conditions, research questions, and contributions.

We next present the results of a post-test to examine whether our theory about inferred urgency could potentially explain why the pre-commitment offer in our field experiment backfired.

Post-test: Initial Evidence on Inferred Urgency. We theorize that when an immediate enrollment option and a pre-commitment option were presented side-by-side in our field experiment (i.e., via simultaneous pre-commitment), employees may have inferred that their university's HR department did not urgently recommend retirement savings. In a pre-registered online experiment (N = 1,499 Mechanical Turk workers), we confirmed that the Simultaneous Pre-commitment mailing from our field experiment conveyed a less urgent recommendation to save than our No Pre-commitment mailing ($p < .001$; see Web Appendix G for more details on this post-test). In the same post-test, we also confirmed that people believed a *Sequential* Pre-commitment offering—that is, sending a second mailing inviting employees to save later only if they neglected an initial mailing with the offer to save immediately—conveyed a more urgent recommendation to save than either our No Pre-commitment mailing ($p = .060$) or the Simultaneous Pre-commitment mailing ($p < .001$).

Together, the findings from this post-test offer tentative support for a mechanism that might explain the results we observed in the field. That is, simultaneous pre-commitment may have reduced immediate adoption of savings and failed to increase overall adoption because the mailing signaled that saving was not urgently recommended. This post-test also reveals that offering *sequential* pre-commitment signals that saving is highly urgent. We next present two large laboratory experiments designed to deductively test our full theory about how simultaneous and sequential pre-commitment may differentially impact the inferences consumers draw about the urgency of marketers' recommendation and, in turn, consumers' farsighted decisions.

STUDY 2: INFERRED URGENCY AND FARSIGHTED DECISIONS

We conducted a pre-registered laboratory experiment in which participants decided (hypothetically) whether and when to enroll in three different benefits programs offered by a new employer. This study tests all four of our hypotheses.

Methods

As pre-registered, we recruited workers on Prolific who were fully employed at a firm other than Prolific and passed one brief attention check. A total of 2,682 participants satisfied these selection criteria and completed the study ($M_{age} = 34.7$, $SD_{age} = 9.1$; 40.3% female).

All participants were asked to imagine that they were offered three benefits programs as a new, full-time employee at Company X: a retirement savings plan, a life insurance program, and a health savings account. They were told that all programs were optional, and money would be deducted from their take-home pay for each program they enrolled in. Participants were required to correctly answer two comprehension check questions before proceeding, and they were allowed to keep trying until they got these comprehension check questions right. Participants were then randomly assigned to one of three conditions: the No Pre-commitment condition, the Simultaneous Pre-commitment condition, or the Sequential Pre-commitment condition.

In the No Pre-commitment condition, participants only had the option to enroll in each benefits program immediately. Specifically, they read: "... if you check the 'Enroll now' box to enroll in a program, Company X will start providing you with the given benefit now and begin deducting from your paycheck as soon as possible.” Then they indicated which program(s) they would enroll in by marking the corresponding checkbox(es).

In the Simultaneous Pre-commitment condition, participants had the option to enroll in each benefits program either immediately or in six months. The instructions explained: “...you can choose ‘Enroll now,’ which means Company X will start providing you with the relevant benefit now and begin deducting from your paycheck as soon as possible. Or you can choose ‘Enroll in 6 months,’ which means Company X will start providing you with the relevant benefit in 6 months and begin deducting from your paycheck in 6 months.” After reading these instructions, participants indicated which program(s) they would enroll in by marking the corresponding checkbox(es). For any programs they selected, they then decided when to enroll (either now or in six months).

In the Sequential Pre-commitment condition, participants were first given the option to enroll in each program immediately, and then for the programs they did not enroll in immediately, they were offered the option to enroll in six months. The condition looked identical to the No Pre-commitment condition through the first (immediate) enrollment decision. Participants who did not immediately enroll in all three programs were next told, “Imagine that the day after you submitted your enrollment decisions, Company X follows up and sends you another online enrollment form.” Participants were then given the option to enroll in the remaining programs in six months by checking the corresponding checkbox(es).

Next, participants in all conditions answered two questions about Company X: “To what extent do you think Company X recommends that employees enroll in the benefits programs as soon as they can?” and “To what extent do you think Company X urgently recommends that employees enroll in the benefits programs” (1 = Not at all; 7 = Very much). We adapted these items from past research on information leakage (McKenzie, Liersch, and Finkelstein 2006) and task urgency (Zhu, Yang, and Hsee 2018) to assess participants' inferences about how urgently Company X recommended they sign up for the benefits programs. The two items were collapsed into a single measure of *inferred urgency* because they hung together well ($r = .70$, $p < .001$).

Finally, we included a set of questions to assess whether decision difficulty could be an alternative mechanism for our predicted results. Specifically, we adapted four items from an existing decision difficulty scale (Goodman et al. 2013): "To what extent [did you find the choice difficult/were you overwhelmed/were you frustrated/were you annoyed] when you were making your enrollment decision?" (1 = Not at all; 7 = Very much). The items hung together well ($\alpha = .85$) and were averaged into one measure of *decision difficulty*. To establish the discriminant validity of these two mechanism measures, we used an exploratory factor analysis and confirmed that the four decision difficulty items loaded on a separate factor than the two urgency items (see Web Appendix H for details). We also checked that the composite score of decision difficulty and that of inferred urgency are only correlated at $r = .15$. At the end of our study, participants were asked about their age, gender, education, and income.

We focus on two pre-registered outcome variables in this study. The first is *immediate adoption,* which measures the number of benefits programs participants elected to enroll in immediately (i.e., by selecting the "Enroll now" option). The second outcome of interest is *overall adoption,* which measures the total number of benefits programs participants elected to enroll in (i.e., by selecting either the "Enroll now" or "Enroll in 6 months" option).

Results

For analyses that include all three conditions, we relied on ordinary least squares (OLS) regressions with heteroskedasticity-robust standard errors where the key predictors are indicators for the Simultaneous Pre-commitment and Sequential Pre-commitment conditions, with the No Pre-commitment condition serving as the reference group. All mediation analyses use 5,000 bootstrapped samples to estimate 95% confidence intervals (CI) around the indirect effects.

Inferred urgency. Providing support for Hypothesis 1, compared to those in the No Pre-commitment condition (M = 4.88, SD = 1.52), we found that participants in the Simultaneous Pre-commitment condition rated Company X's implicit recommendation to enroll in its benefits programs as less urgent (M = 4.55, SD = 1.63; b = -.33, $p < .001$, $d = .21$), whereas participants in the Sequential Pre-commitment condition rated Company X's recommendation to enroll as more urgent (M = 5.22, SD = 1.30; b = .34, $p < .001$, $d = .24$).

Immediate Adoption. Confirming Hypothesis 2 and replicating the results from our field experiment, participants in the Simultaneous Pre-commitment condition signed up for fewer benefits immediately (M = 1.49, SD = .89) than participants in the No Pre-commitment condition (M = 1.86, SD = .78; b = -.37, $p < .001$, $d = .44$). And, consistent with Hypothesis 4, this reduction in immediate adoption was significantly mediated by a drop in the inferred urgency of the recommendation to enroll ($b = -.009$; 95% CI = [-.019, -.001]).

Overall Adoption. Consistent with the results of our field experiment, there was not a statistically significant difference in overall adoption between the Simultaneous Pre-commitment condition (M = 1.84, SD = .81) and the No Pre-commitment condition (M = 1.86, SD = .78; b = -.017, $p = .659$, $d = .02$). Inferred urgency significantly and negatively mediated the relationship between simultaneous pre-commitment and overall adoption (b = -.015, 95% CI = [-.026, -.006]), providing support for Hypothesis 4. These results are in line with our theory that offering people a chance to enroll “now or later” (simultaneous pre-commitment) decreases inferred urgency, which curbs overall adoption of a farsighted behavior.[9]

In support of Hypothesis 3, participants in the Sequential Pre-commitment condition (M = 2.14, SD = .82) signed up for more benefits on average than participants in the No Pre-commitment condition (M = 1.86, SD = .78; b = .28, *p* < .001, *d* = .35). Inferred urgency significantly and *positively* mediated this effect (b = .013, 95% CI = [.003, .025]), suggesting that sequential pre-commitment increases inferred urgency, which predicts greater overall adoption (offering further support for Hypothesis 4).

Using a Wald test, we confirmed that participants in the Sequential Pre-commitment condition enrolled in more benefits on average than participants in the Simultaneous Pre-commitment condition (p < .001, *d* = .37), consistent with Hypothesis 3. And, we confirmed that inferred urgency significantly and positively mediated this difference (b = .019, 95% CI = [.002, .037]), further supporting Hypothesis 4. In other words, compared to the Simultaneous Pre-commitment condition, people may have enrolled in more benefits programs in the Sequential Pre-commitment condition in part because they inferred that adoption was more urgently recommended.

When we added decision difficulty as another potential mediator in all of the aforementioned mediation models, inferred urgency always remained a significant mediator. See detailed results of these multi-mediator models in Web Appendix I.

---

[9] It is possible for simultaneous pre-commitment to have a negative indirect effect on overall adoption via inferred urgency but a null main effect if, as described in our theoretical development, there is a positive indirect effect via a different mechanism (e.g., related to present bias or anticipated resource slack).

Discussion

Study 2 presents support for our complete theory and tests all four of our hypotheses. Offering simultaneous pre-commitment (i.e., an invitation to enroll in benefits “now or later”) decreases the inferred urgency of adopting farsighted behaviors, whereas offering sequential pre-commitment (i.e., an invitation to enroll in benefits “later” only if people don’t enroll “now”) increases inferred urgency (Hypothesis 1). Furthermore, simultaneous pre-commitment decreases immediate adoption of farsighted behaviors (Hypothesis 2), but fails to increase overall adoption (replicating the findings from our field experiment). Meanwhile, sequential pre-commitment increases overall adoption of farsighted behaviors (Hypothesis 3). Importantly, inferred urgency significantly mediates these effects (Hypothesis 4), helping explain the divergent impact of different forms of pre-commitment on consumer choice.[10]

STUDY 3: AN INCENTIVE COMPATIBLE TEST OF OUR FULL THEORY

In Study 3, we sought to conceptually replicate Study 2 with an incentive compatible design in a distinct context. Specifically, we invited people to take a real, 10-minute financial well-being assessment and tested our full theory by examining whether and when they chose to take it.

This study also aims to reconcile our findings with previous literature by measuring an additional mechanism that should make pre-commitment attractive based on extant theory. Specifically, previous work suggests that the immediate costs of adopting a farsighted behavior should feel less aversive when people contemplate taking up the behavior later because they

---

[10] In Web Appendices J and K, we report a pre-registered two-condition version of this study—only containing the Simultaneous Pre-commitment and No Pre-commitment conditions (N = 1,161 MTurk participants). There, we replicated all of the results concerning those two conditions in Study 2. In this study, we also showed that the indirect effect of inferred urgency remained significant after controlling for alternative mechanisms including decision difficulty, confusion, and perceived thoughtfulness.

steeply discount future costs (Frederick, Loewenstein, and O'Donoghue 2002) and expect to have more resources in the future (Zauberman and Lynch 2005). In this sense, taking the financial well-being assessment in our study should feel less costly when pre-commitment is an option, regardless of how it is offered. We operationalized this mechanism with a measure of perceived convenience.

Methods

For this pre-registered study, we recruited 2,398 MTurk participants ($M_{age} = 39.73$, $SD_{age} = 12.06$; 49.0% female) who passed an attention check. To provide a cover story for the purpose of the study, we first asked participants to take a financial knowledge test (Knoll and Houts 2012). Participants next reported their employment status, age, gender, education, and income. Then, we offered them an opportunity to take an optional, unpaid financial well-being assessment. We explained:

> *In collaboration with Dr. [anonymized], a [university affiliation] professor and world expert on financial decision making, we have prepared an assessment that will provide feedback on your financial well-being and offer scientific tips for improving your financial future. Completing the assessment will take about 10 minutes. It is voluntary and won't affect your pay. But we hope that taking the assessment will be worth your time in the long run.*

At this point, participants were randomly assigned to one of three conditions. In the No Pre-commitment condition, participants were invited to take the financial assessment immediately. In the Simultaneous Pre-commitment condition, participants were invited to take the financial assessment and given the choice to either complete it now or in one week. In the Sequential Pre-commitment condition, participants were invited to take the assessment immediately. If they declined, participants were then invited to take the assessment in one week. All participants who chose to take our assessment received a real financial well-being assessment at the time they elected.

After participants made their choice(s), they responded to additional questions. First, we asked participants "To what extent do you think we urgently recommend that you take the financial assessment?" (1 = Not at all; 7 = Very much) to measure *inferred urgency*.

We also asked participants to rate how inconvenient it would be to take the assessment (1 = Not at all inconvenient; 7 = Very inconvenient). We reverse-coded this measure and included "perceived convenience" as a competing mediator in each of the reported mediation models (see Figure 1-4 and Web Appendix L for the full results).

We focus on two pre-registered outcome variables in this study. *Immediate adoption* is a measure of whether a participant elected to take the financial well-being assessment immediately (by selecting the "now" option). *Overall adoption* is a measure of whether a participant ever enrolled (by selecting the "now" option or the "in 1 week" option).

Results

Inferred Urgency. Consistent with Hypothesis 1, participants inferred that our implicit recommendation that they take the assessment was *less* urgent in the Simultaneous Pre-commitment condition (M = 3.57, SD = 1.89; $b = -.30$, $p = .002$, $d = .16$) and *more* urgent in the Sequential Pre-commitment condition (M = 4.23, SD = 1.81; $b = .36$, $p < .001$, $d = .20$) than it was in the No Pre-commitment condition (M = 3.87, SD = 1.79).

Immediate Adoption. Offering some support for Hypothesis 2, participants in the Simultaneous Pre-commitment condition were marginally less likely to immediately enroll in the assessment (28.2 percent) than participants in the No Pre-commitment condition (32.0 percent; $b = -.038$, $p = .098$). In terms of its relative effect, this represents an 11.9% decrease in immediate adoption. In support of Hypothesis 4, this marginal negative effect was mediated by inferred urgency (b = -.011, 95% CI = [-.020, -.004]). This result is consistent with our theory that simultaneous pre-commitment may reduce immediate adoption of farsighted options because it decreases inferred urgency.

Overall Adoption. Offering simultaneous pre-commitment (inviting people to take the assessment "now or later") resulted in greater overall adoption of the assessment (48.0 percent enrolled) than not offering pre-commitment (32.0 percent enrolled; $b = .160$, $p < .001$; representing a 50.0% relative increase). Supporting Hypothesis 4, inferred urgency *negatively* mediated the relationship between simultaneous pre-commitment and overall adoption ($b = -.014$, 95% CI = [-.025, -.005]), as illustrated in Figure 1-4 Panel A. This suggests that offering consumers simultaneous pre-commitment decreases inferred urgency, which may curb take-up of farsighted activities like completing a financial well-being assessment.

Confirming Hypothesis 3, we found that participants in the Sequential Pre-commitment condition were more likely to enroll in the assessment (57.4 percent enrolled) than participants in the No Pre-commitment condition (32.0 percent enrolled; $b = .254$, $p < .001$). In terms of its relative effect, this represents a 79.4% increase in overall adoption. Again, supporting Hypothesis 4, inferred urgency *positively* mediated this effect ($b = .014$, 95% CI = [.006, .023]), as illustrated in Figure 1-4 Panel B. These results are consistent with our theory that sequential pre-commitment may increase overall adoption of a farsighted behavior because it signals heightened urgency.

In addition, a Wald test confirmed that participants in the Sequential Pre-commitment condition were more likely to enroll in the assessment than participants in the Simultaneous Pre-commitment condition ($p < .001$; representing a 19.6% relative increase), which supports Hypothesis 3. And, we confirmed that the difference in overall adoption between the Sequential

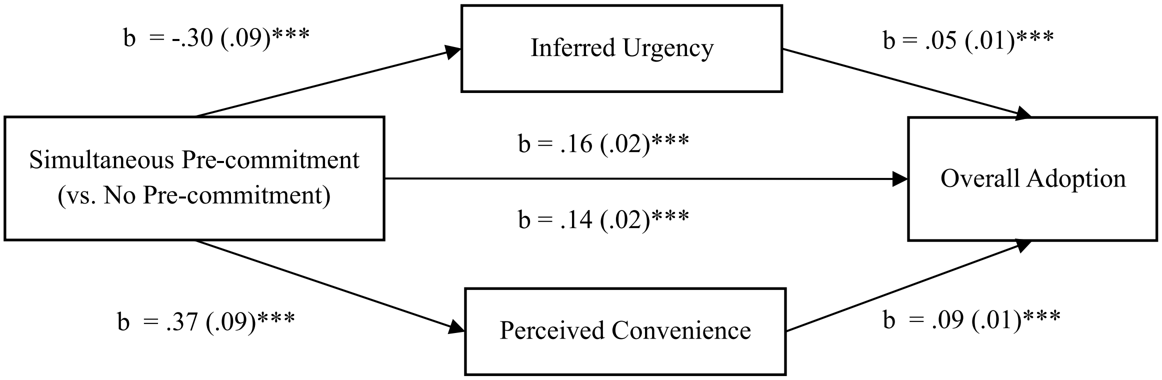
Pre-commitment and Simultaneous Pre-commitment conditions was also mediated by inferred urgency (b = .028, 95% CI = [.018, .040]), in line with Hypothesis 4.

Perceived Convenience. Consistent with predictions from previous research, as shown in Figure 1-4, offering pre-commitment increased the perceived convenience of taking the assessment, regardless of whether the pre-commitment offer was simultaneous (b = .37, p < .001) or sequential (b = .29, p = .001). When participants perceived greater convenience, they were more likely to enroll in the assessment (b = .09, p < .001 for simultaneous pre-commitment and b = .08, p < .001 for sequential pre-commitment). Perceived convenience helps explain why the pre-commitment offers increased enrollment in the assessment (indirect effects: b = .034, 95% CI = [.018, .050] for simultaneous pre-commitment and b = .024, 95% CI = [.010, .040] for sequential pre-commitment). Of particular importance, perceived convenience did not differ between the two pre-commitment designs (p = .36), and thus, it cannot explain why sequential pre-commitment resulted in greater overall adoption than simultaneous pre-commitment (see Web Appendix L for more details).

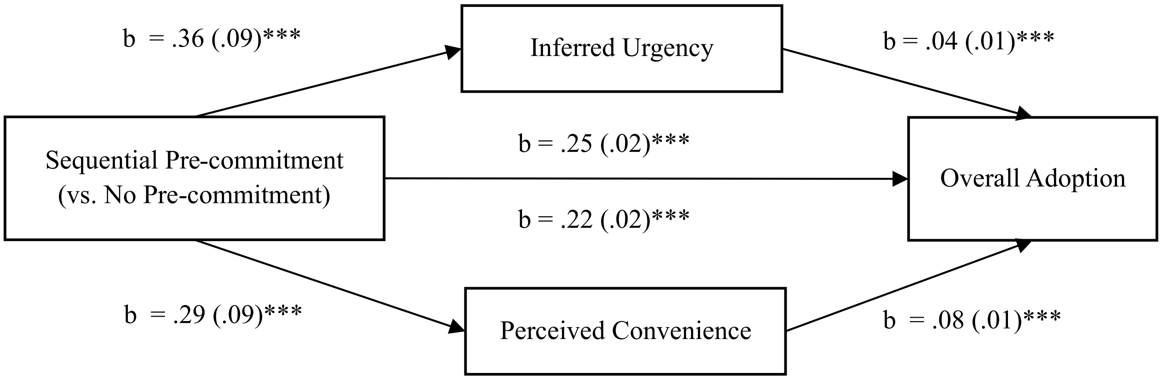
Follow-Through Behavior. Finally, we measured whether participants actually completed the optional financial well-being assessment. Compared to participants in the No Pre-commitment condition (in which 9.6% completed the assessment), participants were more likely to complete the assessment if they were assigned to the Sequential Pre-commitment condition (18.8% completed it; *b* = .092, p < .001) or the Simultaneous Pre-commitment condition (15.5% completed it; *b* = .059, p < .001). A Wald test confirmed that participants in the Sequential Pre-commitment condition were marginally more likely to finish the assessment than participants in the Simultaneous Pre-commitment condition (*p* = .061).

Figure 1-4. Multi-Mediator Models for Overall Adoption (Study 3).
This figure depicts how inferred urgency and perceived convenience explain the effects of offering simultaneous pre-commitment (Panel A) and sequential pre-commitment (Panel B) on overall adoption of the financial well-being assessment. All regression coefficients are unstandardized, and standard errors are presented in parentheses. The coefficients above the paths from Simultaneous Pre-commitment and Sequential Pre-commitment to Overall Adoption represent the total effects, and the coefficients below the paths represent the direct effects. Coefficients significantly different from zero are followed by asterisks (*$p \leq .10$, **$p \leq .05$, *** $p \leq .01$).

Panel A. Simultaneous Pre-commitment



Panel B. Sequential Pre-commitment



Discussion

In an incentive compatible experiment, we again found support for our theory that inferred urgency shapes the way different forms of pre-commitment influence consumer choice. Of note, unlike in our field experiment and in Study 2 where offering simultaneous pre-commitment did not have a significant effect on overall adoption, in this study offering simultaneous pre-commitment increased overall adoption of a financial well-being assessment compared to not offering pre-commitment. This may be because the positive mechanisms that extant theory predict to make pre-commitment offers attractive had a strong influence in this study. Specifically, as shown earlier, participants in the Simultaneous Pre-commitment condition found taking a financial assessment more convenient than participants in the No Pre-commitment condition (see Figure 1-4). Our evidence suggests that the lack of urgency signaled by the simultaneous pre-commitment offer still curbed take-up of the financial well-being assessment, but the added convenience of this offer mattered more to participants. As a result, simultaneous pre-commitment had a *positive* total effect on the take-up of the assessment in this study. This suggests that accounting for the relative magnitudes of various competing mechanisms is necessary in order to predict the main effect of simultaneous pre-commitment on overall adoption of farsighted behaviors. Sequential pre-commitment, however, increases adoption by harnessing both convenience and heightened urgency. Consequently, sequential pre-commitment has an unambiguously positive effect on the overall adoption of farsighted behaviors and produces more adoption of such behaviors than simultaneous pre-commitment.

GENERAL DISCUSSION

Offering pre-commitment—that is, offering people a choice to commit now to do something later—has previously been theorized to uniformly increase the adoption of farsighted behaviors. In this paper, we reexamine this assumption. We develop a theory about how the design of pre-commitment offers can lead consumers to draw inferences about the urgency of a marketer’s recommendation to act, which helps determine when pre-commitment will promote the adoption of farsighted behavior and when it will not.

In a field experiment (Study 1), we show that, contrary to experts' predictions, *simultaneously* offering consumers the option to start saving now and the option to pre-commit to save in the future (i.e., simultaneous pre-commitment) causes people to save less money over our study period than only offering them the option to start saving now. This is because offering simultaneous pre-commitment reduces the number of people who save immediately, without causing more people to save overall. Two pre-registered online experiments (Studies 2 and 3) support our theoretical account of why simultaneous pre-commitment reduces total saving in our field experiment and, conversely, why offering *sequential* pre-commitment has been shown to increase savings (Thaler and Benartzi 2004). Specifically, simultaneous pre-commitment decreases inferred urgency, which helps explain why it reduces immediate adoption and may not increase overall adoption. Moreover, offering sequential pre-commitment—that is, inviting the future adoption of a farsighted behavior only after people decline to adopt it immediately—increases inferred urgency, which helps explain why it increases overall adoption compared to both simultaneous pre-commitment and not offering pre-commitment. Altogether, across one field and two laboratory experiments including over 10,000 participants, involving diverse populations, and examining a variety of farsighted behaviors, we provide robust support for our theory about the role of inferred urgency in consumers' responses to different types of pre-commitment offerings.

Practical Implications

The current research has important implications for marketers and policy makers hoping to increase the adoption of farsighted behaviors like saving. First, our research sheds light on how to design effective pre-commitment strategies. Prior research has primarily shown that offering people a single pre-commitment option (i.e., inviting people to commit now to adopt a behavior later) leads to more future-oriented choices than only inviting people to adopt the behavior now. But when applying this knowledge, marketers and policy makers may assume that they can simply add a pre-commitment option on top of an immediate adoption option (perhaps due to their assumption that offering more options can better cater to individuals' heterogeneous preferences). They may particularly favor simultaneous pre-commitment because sequential pre-commitment can be costlier to implement (given that it requires repeated communication). These same considerations motivated us to test the efficacy of simultaneous pre-commitment in our field experiment. However, our work suggests that simultaneous pre-commitment is less effective than sequential pre-commitment and sometimes (as in our field experiment) less effective than only offering the option to adopt a behavior immediately. Importantly, it can lead consumers to delay action in contexts where such delays are costly (e.g., delays to enroll in savings programs can lead to less accumulated savings, delays to update software can increase likelihood of malware attacks, delays to obtain recommended health screenings can prevent early disease detection, and delays to vaccinate can lead to unnecessary illness).

Furthermore, our findings highlight the value of understanding when choice sets presented by marketers and policy makers inadvertently communicate an urgent recommendation (or a lack thereof) to take action. More generally, our research suggests that seemingly innocuous aspects of the design of interventions (e.g., the simultaneous vs. sequential presentation of a pre-commitment option) can shape people's inferences and responses. As marketers attempt to leverage psychological principles (e.g., present bias) to motivate behavior change and adopt interventions from prior research, it is natural to modify the designs of those interventions to fit specific field settings, but these modifications can inadvertently leak information that ultimately harms the efficacy of the interventions. This points to the critical need to pilot-test interventions and probe what inferences they produce (Reiff et al. 2022). After such pilots, marketers and policy makers can then revise their designs to guard against unintentional information leakage before rolling out interventions at scale.

Limitations and Future Directions

Our research has several limitations, which also open up interesting directions for future research. First, our studies find that inferred urgency *partially* mediates the relationship between pre-commitment offers and farsighted decisions, suggesting that additional processes beyond those we theorized about in this paper may influence people's responses to pre-commitment offers. For instance, participants may choose randomly over options in a given menu, which could have contributed to the observed effects of the pre-commitment offers in our hypothetical scenario in Study 2. Since simultaneous pre-commitment is the only offer we studied that included "now" and "later" enrollment options on the same menu, random choice could partially explain why people in the simultaneous pre-commitment condition were less likely to enroll immediately than those only given the option to enroll "now" in the control condition.

Another limitation of this research is that sequential pre-commitment is the only offer we studied that asks people *twice* whether they would adopt a behavior, and the mere repetition could contribute to the offer's positive effects on overall take-up of farsighted behaviors.[11] Future research should further explore these alternative accounts and others.

We also study farsighted behaviors involving both monetary (Studies 1 and 2) and time (Study 3) costs, and we find support for our theory about inferred urgency across both resource

---

[11] To test this account, sequential pre-commitment could be compared to an additional condition that first asks people whether they would like to adopt a behavior immediately, and then if they decline, asks them to consider immediate adoption again (e.g., "Are you sure about not increasing your savings now?"). Existing theory and the results of Study 3 suggest that offering people the immediate enrollment option twice may be less effective than sequential pre-commitment at increasing adoption of farsighted behaviors because the former does not include a pre-commitment offer and cannot leverage the psychology that people feel less averse to adopting farsighted behaviors in the future.

types. That said, the effect of simultaneous pre-commitment on overall adoption seemed to vary across resource types; simultaneous pre-commitment had a null effect in Studies 1 and 2 but a positive effect in Study 3. As we proposed earlier, the impact of simultaneous pre-commitment on overall adoption appears to depend on the relative magnitude of competing mechanisms: a lack of inferred urgency reduces adoption, while mechanisms related to present bias and resource slack increase adoption. The latter positive mechanisms may have played a stronger role in driving the impact of simultaneous pre-commitment when time costs (as opposed to monetary costs) were involved.[12] Future research should further investigate how the overall impact of pre-commitment offers varies with the resource type.

Further, all of the pre-commitment offers we study involve *non-binding* commitments; that is, after committing to do something later (e.g., to increase their savings contribution rate in six months), people can always change their mind and nullify their decisions. An alternative design for pre-commitment offers could include *binding* commitments, which require people to stick to their initial commitment, and this design feature may be a key determinant of take-up (Karlan and Linden 2015). Future research should test whether our theory about inferred urgency also applies to binding pre-commitment offers.

In addition, though our research suggests that simultaneous pre-commitment leads consumers to infer that the adoption of a farsighted behavior is not urgently recommended, marketers and policy makers may be able to improve the effectiveness of simultaneous pre-commitment offers by changing the framing of options. For instance, when Google prompts consumers to update their notification settings, they offer simultaneous pre-commitment,

---

[12] We speculate that this may be because people typically expect to have more discretionary time in the future than in the present but expect less growth when thinking about their discretionary money (Zauberman and Lynch 2005). Thus, pre-commitment options that allow people to delay time costs into the future–when they think they will have more time to spend–will be particularly attractive (more so than pre-commitment options that delay monetary costs).

presenting the options “continue” or “ask me later.” However, Google prints “continue” in bright blue letters, while “ask me later” appears in a light grey font, which may signal that the company recommends completing the update sooner rather than later. Future research testing different strategies for changing the information leaked by pre-commitment offers would be valuable. Alternatively, Beshears et al. (2021) suggest leveraging a conceptually distinct psychological process to improve simultaneous pre-commitment. They show that offering pre-commitment with the delayed behavior starting shortly after a moment that feels like a fresh start (e.g., after a birthday, the first day of spring) increases retirement savings (relative to offering pre-commitment at an equivalent, unlabeled time delay; e.g., in 2 months). Future work can more broadly examine how to frame pre-commitment options in ways that enhance overall adoption of farsighted behaviors.

Finally, it would be valuable to study moderators of the effects documented in our research. For instance, the extent to which people are influenced by the urgency of a marketer’s recommendation may depend on consumers’ trust in that marketer. Future research is needed to understand how underlying attitudes towards whoever presents choices may moderate responses to implicit recommendations and thus influence the effects of different types of pre-commitment offerings.

CHAPTER 1

APPENDIX

Web Appendix A. Expert Prediction Surveys

Panel A. Survey with Marketing Professors

Methods. We emailed all 408 tenure-track marketing professors at the top 30 U.S. business schools (as ranked by *U.S. News and World Report*) on November 15, 2017 and invited them to take a short survey. The survey asked professors to imagine that a company was testing two campaigns with the goal of increasing employees' contributions to its retirement saving plan. The first campaign offered employees the option to start saving now (i.e., the No Pre-commitment campaign) and the second mailing gave employees the choice to either (i) start saving now or (ii) pre-commit to begin saving in four months (i.e., the Simultaneous Pre-commitment campaign). Note that we did not label these campaigns this way in the survey. The survey asked professors to predict which mailing would lead employees to save more over nine months.[13] Demographic information was not collected on respondents. As pre-registered, we stopped data collection after seven days and excluded five professors who reported that they were familiar with our Study 1 results, leaving us with 101 professors for analysis.

Results. Of the 85 respondents who clearly favored one campaign over the other (rather than predicting the two campaigns would yield the same results or selecting "I don't know"), the proportion who believed the Simultaneous Pre-commitment campaign would lead to more savings (62.4%) was significantly higher than 50% (p = .023 in a two-sided, one-sample proportion test). As a more conservative test, we also analyzed the whole sample of 101 professors and treated their responses as a ternary outcome variable, taking the following three values: "Simultaneous Pre-commitment" when the prediction was that the Simultaneous Pre-commitment campaign would lead to more savings, "No Pre-commitment" when the prediction was that the No Pre-commitment campaign would lead to more savings, or "other" when the prediction was that the two campaigns would lead to the same amount of savings or the response was "I don't know". We then calculated the p-value for the following null hypothesis: the proportion of the population responding Simultaneous Pre-commitment = the proportion of the population responding No Pre-commitment = some fixed number x. We chose x to generate the most conservative p-value by setting x equal to half of the sample proportion who responded either Simultaneous Pre-commitment or No Pre-commitment (i.e., x = .5 * (53 + 32) / 101 = .421). Under the null hypothesis with x = .421, we calculated the probability of seeing a proportion of Simultaneous Pre-commitment responses (out of Simultaneous Pre-commitment plus No Pre-commitment responses) that is as extreme or more extreme in the two-sided sense than our observed probability (i.e., 53 / (53 + 32) = .624). This probability (i.e., the p-value for this test) is .041.

Panel B. Survey with Financial Advisors

---

[13] We think "save more over the nine months" is likely interpreted by survey respondents as the accumulated savings during that time window, which corresponds closely with the average savings rate variable we used in our field experiment as an outcome measure.

Methods. We recruited financial advisors specialized in retirement savings at (1) a symposium organized by Voya Financial on September 27, 2017, and (2) an event organized by LPL Financial on October 2, 2017. In the middle of each event, consultants were presented with a slide describing two savings campaigns—a campaign that only presented people with the option to start saving now (i.e., the No Pre-commitment campaign) and a campaign that allowed people to either start saving now or pre-commit to begin saving in four months (i.e., the Simultaneous Pre-commitment campaign). Note that we did not label these campaigns this way on the slide. We then invited the audience to take an anonymous poll. Financial advisors were asked to predict which campaign would lead to more savings over the following nine months. Demographic information on respondents was not collected in this one-question poll. A total of 239 financial advisors responded to the poll.

Results. Of the 229 respondents who favored one campaign over the other (as opposed to predicting the two campaigns would yield the same results or selecting "I don't know"), the proportion who believed the Simultaneous Pre-commitment campaign would lead to more savings (73.8%) was significantly higher than 50% ($p < .001$ in a two-sided, one-sample proportion test). We conducted the same conservative analysis on the whole sample of 239 financial advisors as we did to marketing professors. Here, x again equals to half of the sample proportion who responded either Simultaneous Pre-commitment or No Pre-commitment. That is, $x = .5 * (169 + 60) / 239 = .479$. Under the null hypothesis with $x = .479$, we calculated the probability of seeing a proportion of Simultaneous Pre-commitment responses (out of Simultaneous Pre-commitment plus No Pre-commitment responses) that is as extreme or more extreme in the two-sided sense than our observed probability (i.e., 169 / (169 + 60) = .738). This probability (i.e., the p-value for this test) is less than .001.

Web Appendix B. Additional Information on Sample and Randomization (Study 1)

Randomization

Employees were randomized into one of three primary mailing conditions: the No Pre-commitment mailing, the Simultaneous Pre-commitment mailing, and the Framed Simultaneous Pre-commitment mailing.[14] Employees assigned to the No Pre-commitment mailing were encouraged to sign up to save (or to save more) immediately. Those assigned to the Simultaneous Pre-commitment mailing were given the opportunity to sign up to save (or to save more) either immediately or after a time delay ranging from two to six months (e.g., "in two months"). Finally, those assigned to the Framed Simultaneous Pre-commitment mailing received a mailing identical to the Simultaneous Pre-commitment mailing, except that the pre-commitment time window referenced (e.g., "in two months") was replaced by a reference to a temporal landmark associated with the same length of delay (e.g., "following your next birthday," "following Thanksgiving"). The temporal landmarks were either holidays (Thanksgiving, New Year's, Martin Luther King Day, Valentine's Day, and the Spring Equinox) or employees' birthdays. Some of these temporal landmarks feel like the beginning of a new cycle and are associated with fresh starts (New Year's, the first day of spring, employees' birthdays), and some are ordinary holidays that do not signal a new beginning (Thanksgiving, Martin Luther King Day, Valentine's Day; Beshears et al. 2021). Note that in the current paper we only used data for employees

---

[14] This condition is labelled in the manuscript as "the Simultaneous Pre-commitment Condition Linked with Temporal Landmarks."

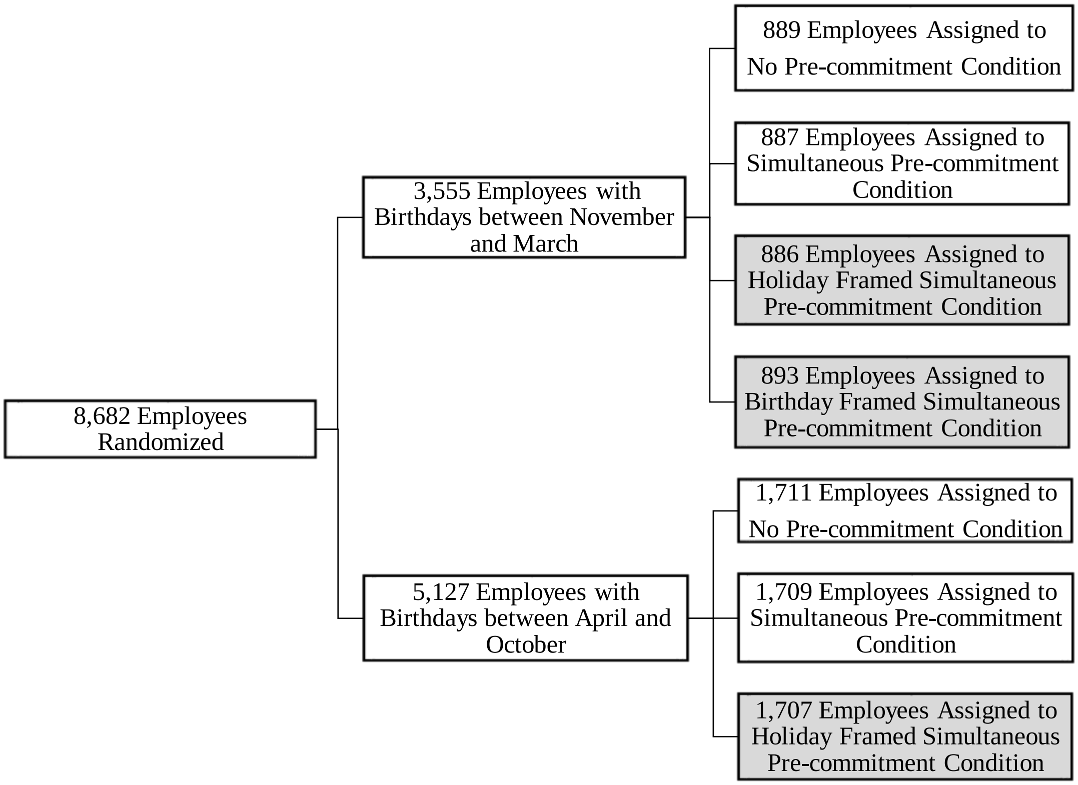
assigned to receive either the No Pre-commitment mailing or the Simultaneous Pre-commitment mailing.

As explained in the manuscript, randomization was stratified by birth month within each university because birthdates partially determined which mailings employees received. Specifically, as illustrated in Figure A1-1, employees were divided into two sub-groups before randomization: those with birthdays between November and March (hereafter referred to as the "November-March birthday group"), and others (hereafter referred to as the "April-October birthday group"). Those with birthdays between November and March were divided evenly among four experimental mailings: the No Pre-commitment mailing, the Simultaneous Pre-commitment mailing, and two sub-categories of the Framed Simultaneous Pre-commitment mailings—the Birthday Framed Simultaneous Pre-commitment mailing that offered employees an opportunity to begin saving (or to save more) following their next birthday and the Holiday Framed Simultaneous Pre-commitment mailing that offered employees an opportunity to begin saving (or to save more) following a future holiday. Among employees in the November-March birthday group that were offered pre-commitment, the length of time until the offered pre-commitment option would take effect was determined by employees' birth month. For instance, consider an employee whose birthday is in December. Recall that our mailings went out in early October. If this employee were assigned to the Simultaneous Pre-commitment mailing, she would be offered the opportunity to start saving "in three months" (i.e., in January). If she were assigned to the Holiday Framed Simultaneous Pre-commitment mailing, she would be offered the opportunity to start saving "after New Year's" (i.e., in January). If she were assigned to the Birthday Framed Simultaneous Pre-commitment mailing, she would be offered the opportunity to start saving "after her next birthday" (again, in January).

Employees with birthdays between April and October were divided evenly among receiving the No Pre-commitment mailing, the Simultaneous Pre-commitment mailing, and the Holiday Framed Simultaneous Pre-commitment mailing. Among employees in the April-October birthday group who were offered pre-commitment, the length of time until the offered pre-commitment would take effect was randomized to be from two to six months (in the Simultaneous Pre-commitment condition) or from Thanksgiving to the Spring Equinox (in the Holiday Framed Simultaneous Pre-commitment condition). Every employee in the Holiday Framed Simultaneous Pre-commitment condition was yoked with an employee in the (un-framed) Simultaneous Pre-commitment condition who was offered the opportunity to start saving (or to save more) at the same time delay. For example, an employee who was randomly assigned to have the opportunity to begin saving after New Year's was yoked with an employee who had the opportunity to begin saving in three months. Notably, past research suggests that New Year's is a particularly meaningful fresh start opportunity (Dai et al. 2014). Thus, among employees in the Holiday Framed Simultaneous Pre-commitment condition, we oversampled assignment to the sub-group that allowed employees to pre-commit to begin saving (or to save more) "after New Year's". Correspondingly, we oversampled assignment to the sub-group of the (un-framed) Simultaneous Pre-commitment condition that allowed employees to pre-commit to begin saving (or to save more) "in three months."

Note that for the current study we only used data from employees that were assigned to the No Pre-commitment or Simultaneous Pre-commitment conditions (889 + 887 + 1,711 + 1,709 = 5,196).

Figure A1-1. Randomization.
This figure shows how employees were assigned to conditions. Note the gray boxes reflect the Framed Simultaneous Pre-commitment condition, which are reported in Beshears et al. (2021) and not included in the current paper.



There were 42 employees that were randomized but not included in the analysis because they did not have data collected, were terminated before the baseline data collection, or had conflicting dates of birth.

Balance Checks

Table A1-1. Summary Statistics by Condition in Study 1.
This table summarizes the mean and standard deviation (in parentheses) of key control variables used in our analyses by experimental condition. The last column shows p-values from statistical tests comparing the conditions.

	No Pre-commitment Mailing	**Simultaneous Pre-commitment Mailing**	**Simultaneous Pre-commitment Vs. No Pre-commitment**
			p-value
Female	52.65%	51.85%	.562
Age (years)	43.20 (12.32)	43.00 (11.77)	.548
Tenure (years)	9.51 (9.14)	9.54 (8.93)	.906
Baseline Salary ($USD)	56,505.19 (35,234.21)	58,505.26 (36,111.88)	.043
Faculty	11.62%	12.75%	.211

Web Appendix C. Descriptions of Non-Targeted Plans (Study 1)

Table A1-2. Non-Targeted Plans.

The experimental mailings encouraged employees to increase contributions to a targeted plan (see Table 3 in the manuscript). This table describes other retirement plans (i.e., “Non-targeted plans") that were also available to employees.

University	Plan	Eligibility	Employee Contributions	Employer Contributions	Automatic Enrollment
A	*Plan 1*	Determined based on employee's position and scheduled hours of service	None	The University pays the full cost by contributing 10% of the employee’s base pay. The base pay limit was $255,000 for 2013 and $260,000 for 2014.	No
B	*Plan 1*	Regular or fixed-term employees scheduled to work at least 1,000 hours per fiscal year and not currently actively participating in Plan 2	1% of the employee’s eligible gross earnings on a pre-tax basis	Matched by an 8% contribution rate from the University	No
	Plan 2	Regular or fixed-term employees who were hired prior to June 30, 1993 and scheduled to work at least 20 hours per week for a minimum of 720 hours per fiscal year	1% of the employee’s eligible gross earnings on an after-tax basis	There are several benefit calculation formulas. This plan uses the formula that maximizes employee benefits.	No
C	*Plan 1*	All faculty and staff member in a benefits-eligible title who are age 21 or older	None	*Employer contribution rates when the employee's annual salary is below the Social Security Wage Base:* i) 7% (employee age < 50) ii) 10% (employee age ≥ 50) *Employer contribution rates when the employee's annual salary is above the Social Security Wage Base:* i) 12% (employee age < 50) ii) 15% (employee age ≥ 50)	Yes
	Plan 2	All employees who earn at least 140% of the Social Security Wage Base	An elected percentage of the employee’s eligible earnings on a pre-tax basis	None	No
D	*Plan 1*	All employees except for student workers, hospital employees, leased employees, and those in post-doctoral positions	An elected percentage of the employee’s eligible earnings on a pre-tax basis	None	No
	Plan 2	Certain employees who, as of July 2000, worked at least 1,000 hours per year and opted not to be covered by the targeted plan	None	*Monthly defined benefit payment:* (Final average pay[1] * Years of participation in Plan 2 * 1.25%)/12	Yes

[1] Final average pay equals average pay for the five years of highest pay that fall within the last ten years of plan participation.

Notes: Table A1-1 is identical to the one presented in Beshears et al. (2021) because it relies on the same retirements plans in the same universities.

Web Appendix D. Summary of Standard Practices in Retirement Savings Research

We examined all of the papers covered in the most comprehensive review of the literature on contributions to pension plans in recent years (Choi 2015). We identified 33 empirical papers that analyzed contributions to pension plans using savings data in the field.

We coded the outcome variables that these papers analyzed as well as the regression models that these papers used to predict their outcome variables (see Table A1-3 for a summary). The most common outcome variable studied was whether individuals participated in a given savings plan (n = 21 papers). Among the papers that examined this outcome variable, 16 papers (76%) used OLS regressions to predict such a binary outcome variable.

The second most common outcome variable studied in these papers was people's average pension plan contribution rate (n = 13 papers). Among the papers that examined this outcome variable, 12 papers (92%) used OLS regressions to predict this outcome variable.

We also coded the control variables that these papers included in their regressions. It is a standard practice to control for employees' demographics (e.g., age, gender), income, and employer characteristics. For example, over 80% of the empirical papers in this survey controlled for age and salary. Among the six papers that reported on *field experiments* testing the efficacy of an intervention in increasing savings behavior, four papers (67%) controlled for age and five papers (83%) controlled for salary.

Table A1-3. Summary of 33 Empirical Papers on Pension Savings Reviewed in Choi (2015)

	Field	Dependent Variable:		Controls:	
Citation	Experiment	Contribution Rate	Savings Plan Participation	Age	Income
Ameriks J et al. (2003)				Yes	Yes
Argento et al. (2015)				Yes	Yes
Bassett et al. (1998)			Yes (model: OLS)	Yes	Yes
Bayer et al. (2009)		Yes (model: OLS)	Yes (model: OLS)		
Benartzi et al. (2012)			Yes (model: OLS)	Yes	Yes
Bernheim and Garrett (2003)		Yes (model: OLS)	Yes	Yes	Yes
Beshears et al. (2013)		Yes (model: OLS)		Yes	Yes
Beshears et al. (2015)	Yes	Yes (model: OLS)	Yes (model: OLS)	Yes	Yes
Brown JR et al. (2011)				Yes	Yes
Card and Ransom (2011)		Yes		Yes	Yes
Carroll et al. (2009)		Yes (model: OLS)		Yes	Yes
Chetty et al. (2014)		Yes (model: OLS)		Yes	Yes
Choi et al. (2002)			Yes (model: OLS)	Yes	
Choi et al. (2004)		Yes (model: OLS)	Yes (model: OLS)	Yes	Yes
Choi et al. (2009)		Yes (model: OLS)		Yes	Yes
Choi et al. (2010)	Yes			Yes	Yes
Choi et al. (2011)			Yes	Yes	Yes
Choi et al. (2012)	Yes	Yes (model: OLS)			
Duflo and Saez (2002)			Yes (model: OLS)	Yes	Yes
Duflo and Saez (2003)	Yes		Yes (model: OLS)	Yes	Yes
Duflo et al. (2006)	Yes		Yes (model: OLS)		Yes
Dworak-Fisher (2011)			Yes	Yes	Yes
Engelhardt and Kumar (2007)			Yes	Yes	Yes
Even and Macpherson (2005)			Yes	Yes	Yes
Goda et al. (2014)	Yes			Yes	Yes
Huberman et al. (2007)		Yes (model: OLS)	Yes (model: OLS)	Yes	Yes
Lusardi and Mitchell (2007)				Yes	Yes
Madrian and Shea (2001)		Yes (model: OLS)	Yes (model: OLS)	Yes	Yes
Mitchell et al. (2007)		Yes (model: OLS)	Yes (model: OLS)	Yes	Yes
Papke (1995)			Yes (model: OLS)		
Papke and Poterba (1995)			Yes (model: OLS)		
Sethi-Iyengar et al. (2004)			Yes (model: OLS)	Yes	Yes
Skimmyhorn (2013)			Yes (model: OLS)	Yes	Yes

References

1. Ameriks J, Caplin A, Leahy J (2003) Wealth accumulation and the propensity to plan. *Quarterly Journal of Economics* 118:1007–46
2. Argento R, Bryant VL, Sabelhaus J (2015) Early withdrawals from retirement accounts during the Great Recssion *Contemporary Economic Policy* 33:1-16
3. Bassett WF, Fleming MJ, Rodrigues AP (1998) How workers use 401(k) plans: The participation, contribution, and withdrawal decisions. *National Tax Journal* 51:263-89
4. Bayer PJ, Bernheim BD, Scholz JK (2009) The effects of financial education in the work-place: Evidence from a survey of employers. *Economic Inquiry* 47:605-24
5. Benartzi S, Ehud P, Thaler RH (2012) Choice architecture and retirement saving plans. E. Shafir, ed. *The Behavioral Foundations of Policy* (Russell Sage Foundation and Princeton University Press, Princeton, NJ), 245-263.
6. Bernheim BD, Garrett DM (2003) The effects of financial education in the workplace: evidence from a survey of households. *Journal of Public Economics* 87:1487–519
7. Beshears J, Choi JJ, Laibson D, Madrian BC (2013) Simplification and saving. *Journal of Economic Behavior and Organization* 95:130-145.
8. Beshears J, Choi JJ, Laibson D, Madrian BC (2015) The effect of providing peer information on retirement savings decisions. *Journal of Finance* 70: 1161-1201
9. Brown JR, Farrell AM, Weisbenner SJ (2011) The downside of defaults. *Working Paper, University of Illinois Urbana-Champaign*
10. Card D, Ransom M (2011) Pension plan characteristics and framing effects in employee savings behavior. *Review of Economics and Statistics* 93:228-243
11. Carroll GD, Choi JJ, Laibson D, Madrian BC, Metrick A (2009) Optimal defaults and active decisions. *Quarterly Journal of Economics* 124:1639-74
12. Chetty R, Friedman JN, Leth-Petersen S, Nielsen TH, Olsen T (2014) Active vs. passive decisions and crowd-out in retirement savings accounts: Evidence from Denmark. *Quarterly Journal of Economics* 129: 1141-1219
13. Choi JJ, Haisley E, Kurkoski J, Massey C (2012) Small cues change savings choices. *Working Paper, National Bureau of Economic Research*
14. Choi JJ, Laibson D, Madrian BC (2010) $100 bills on the sidewalk: Suboptimal investment in 401(k) plans. *Review of Economics and Statistics* 93:748-63
15. Choi JJ, Laibson D, Madrian BC (2011) Reducing the complexity costs of 401(k) participation through Quick Enrollment™. Wise DA, ed. *Developments in the Economics of Aging*, (University of Chicago Press, Chicago, IL), 57-82
16. Choi JJ, Laibson D, Madrian BC, Metrick A (2002) Defined contribution pensions: Plan rules, participant decisions, and the path of least resistance. Poterba JM, ed. *Tax Policy and the Economy, Volume 16* (MIT Press, Cambridge, MA), 67-114
17. Choi JJ, Laibson D, Madrian BC, Metrick A (2004) For better or for worse: Default effects and 401(k) savings behavior. Wise DA, ed. *Developments in the Economics of Aging*, (University of Chicago Press, Chicago, IL), 81-121
18. Choi JJ, Laibson D, Madrian BC, Metrick A (2009) Reinforcement learning and savings behavior. *Journal of Finance* 64:2515-34

19. Duflo E, Gale W, Liebman J, Orszag P, Saez E (2006) Saving incentives for low- and middle-income families: Evidence from a field experiment with H&R Block. *Quarterly Journal of Economics* 121:1311-46
20. Duflo E, Saez E. (2002) Participation and investment decisions in a retirement plan: The influence of colleagues' choices. *Journal of Public Economics* 85:121-48
21. Duflo E, Saez E. (2003) The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *Quarterly Journal of Economics* 118:815-42
22. Dworak-Fisher K (2011) Matching matters in 401(k) plan participation. *Industrial Relations* 50:713-37
23. Engelhardt GV, Kumar A. (2007) Employer matching and 401(k) saving: Evidence from the health and retirement study. *Journal of Public Economics* 91:1920-43
24. Even WE, Macpherson DA (2005) The effects of employer matching in 401(k) plans *Industrial Relations* 44:525-49
25. Goda GS, Manchester CF, Sojourner A (2014) What will my account really be worth? Experimental evidence on how retirement income projections affect saving. *Journal of Public Economics* 119:80-92
26. Huberman G, Iyengar SS, Jiang W (2007) Defined contribution pension plans: Determinants of participation and contributions rates. *Journal of Financial Services Research* 31:1-32
27. Lusardi A, Mitchell OS (2007) Baby Boomer retirement security: The roles of planning, financial literacy, and housing wealth. *Journal of Monetary Economics* 54:205-24
28. Madrian BC, Shea DF (2001) The power of suggestion: Inertia in 401(k) participation and savings behavior. *Quarterly Journal of Economics* 116:1149-87
29. Mitchell OS, Utkus SP, Yang T (2007) Turning workers into savers? Incentives, liquidity, and choice in 401(k) plan design. *National Tax Journal* 60:469-89
30. Papke LE (1995) Participation in and contributions to 401(k) pension plans: Evidence from plan data. *The Journal of Human Resources* 30:311-25
31. Papke LE, Poterba JM (1995) Survey evidence on employer match rates and employee saving behavior in 401(k) plans. *Economics Letters* 49:313-17
32. Sethi-Iyengar S, Huberman G, Jiang W (2004) How much choice is too much? Contributions to 401(k) retirement plans. Mitchell OS, Utkus SP, ed. *Pension Design and Structure: New Lessons from Behavioral Finance* (Oxford University Press, Oxford), 83-95
33. Skimmyhorn W (2013) Assessing financial education: Evidence from a personal financial management course. *Working Paper, United States Military Academy West Point*

Web Appendix E. Additional Analyses and Robustness Checks (Study 1)

Panel A. Additional Analyses

Analysis of enrollment in all plans (Table A1-4). In the manuscript we focus on changes to contribution rates in the targeted plan because the mailings containing the experimental manipulation encouraged enrollment in the targeted plans. However, employees could have also changed their contribution rates to non-targeted plans that were also offered by their employer (see Web Appendix C for details). In particular, if employees increased contributions to the targeted plan in response to our mailing by simply shifting contributions away from non-targeted plans, this would not reflect an actual increase in savings. Thus, here we examine when and how contributions to *all* plans—including targeted and non-targeted plans—changed, which allows us to more comprehensively capture the effects of offering simultaneous pre-commitment on retirement savings decisions. The table below uses the primary regression specification, examining effects of the Simultaneous Pre-commitment mailing on immediate adoption, overall adoption, and average savings rate in all plans. Importantly, we obtained qualitatively similar results for all plans and targeted plans.

Table A1-4. Analysis of Enrollment in All Plans

	Model 1: Immediate Adoption	Model 2: Overall Adoption	Model 3: Average Savings Rate
Simultaneous Pre-commitment	-.018**	-.012	-.0027***
	(.007)	(.009)	(.001)
R-squared	.09	.12	.52
Observations	5,196	5,196	5,196

Standard errors robust to heteroskedasticity in parentheses. The regressions control for gender, age decile, tenure decile, salary decile, faculty status, and birth month. We allowed the coefficients on the control variables to vary by university. **p* ≤ .10, ***p* ≤ .05, *** *p* ≤ .01.

Contribution rate increase (Table A1-5). For employees who increased their contribution rates from September 2013 (right before our experiment) to June 2014 (the last month of the observation period), we measured the amount of increase in their contribution rate by taking the difference in their contribution rates (as revealed in the payroll data) between September 2013 and June 2014. We refer to this difference as "contribution rate increase." As explained earlier, all the mailings offered a default contribution rate of either 3% or 5% which employees could not change via mail, and only University D sent mailings to employees who were already enrolling at a non-zero contribution rate. As a result, among employees who increased their contribution rates to the targeted plans, approximately 50% increased by 3% or 5%. Notwithstanding, some employees set their own contribution rates in the targeted plans via phone or internet. Thus, we could assess whether the Simultaneous Pre-commitment mailing had a separate impact on the size of the increase in contribution rates, independent of its effect on immediate or overall adoption. When we used the OLS regression specification described in the paper to predict contribution rate increase, we did not find evidence that the size of contribution rate changes differed statistically significantly between conditions.

Table A1-5. Contribution Rate Increase

	Model 1: Contribution Rate Increase
Simultaneous Pre-commitment	-.005
	(.005)
R-squared	.40
Observations	621

Standard errors robust to heteroskedasticity in parentheses. The regressions control for gender, age decile, tenure decile, salary decile, faculty status, and birth month. We allowed the coefficients on the control variables to vary by university.
*$p \leq .10$, **$p \leq .05$, *** $p \leq .01$.

Effects by university (Tables A1-6 and A1-7). We added interactions between each university and the Simultaneous Pre-commitment indicator to our primary regression specification. In Table A1-6, we report the coefficients on the interaction terms, and in Table A1-7 we report the simple effects of simultaneous pre-commitment in each university (estimated with linear combinations of the coefficients from the regressions in Table A1-6). Note that some employees at University D were already enrolled in the targeted plan, and they were asked to increase their contribution rates in the experimental mailings. We analyzed their responses (*University D (Already Enrolled)*) separately from the rest of University D's employees in this experiment who were not yet enrolled (*University D (Not Yet Enrolled)*).

Table A1-6. Interactions by University

	Model 1: Immediate Adoption	Model 2: Overall Adoption	Model 3: Average Savings Rate
Simultaneous Pre-commitment	-.009 (.014)	-.002 (.019)	.0007 (.001)
*Simultaneous Pre-commitment * University B*	-.009 (.024)	-.009 (.029)	-.0003 (.002)
*Simultaneous Pre-commitment * University C*	-.005 (.017)	.008 (.023)	-.0031* (.002)
*Simultaneous Pre-commitment * University D (Not Yet Enrolled)*	-.012 (.022)	-.025 (.029)	-.0029** (.001)
*Simultaneous Pre-commitment * University D (Already Enrolled)*	-.041 (.034)	-.044 (.042)	-.0018 (.002)
R-squared	.07	.09	.11
Observations	5,196	5,196	5,196

Standard errors robust to heteroskedasticity in parentheses. The regressions control for gender, age decile, tenure decile, salary decile, faculty status, and birth month. We allowed the coefficients on the control variables to vary by university. *$p \leq .10$, **$p \leq .05$, *** $p \leq .01$.

Table A1-7. Simple Effects of Simultaneous Pre-commitment in Each University

	Model 1: Immediate Adoption	Model 2: Overall Adoption	Model 3: Average Savings Rate
Simultaneous Pre-commitment in University A	-.009	-.002	.0007
	(.014)	(.019)	(.001)
Simultaneous Pre-commitment in University B	-.019	-.011	.0005
	(.019)	(.022)	(.002)
Simultaneous Pre-commitment in University C	-.014	.006	-.0024*
	(.010)	(.013)	(.001)
Simultaneous Pre-commitment in University D (Not Yet Enrolled)	-.021	-.026	-.0022**
	(.017)	(.021)	(.001)
Simultaneous Pre-commitment in University D (Already Enrolled)	-.050	-.045	-.0010
	(.030)	(.038)	(.001)

Standard errors robust to heteroskedasticity in parentheses. $*p \leq .10$, $**p \leq .05$, $*** p \leq .01$.

Panel B. Robustness Checks

Measuring decisions with cards mailed back to the one university that tracked responses (Table A1-8). Only one of the four universities provided us with data indicating which employees returned response postcards and what enrollment decisions they made on those cards. Using the mailing response data from this university, we created two outcome variables. "Immediate adoption" equaled one if an employee elected to *immediately* increase their contribution rate to the targeted plan by marking the immediate option on the response card, and zero otherwise. "Overall adoption" equaled one if an employee elected to increase their contribution rate to the targeted plan by marking either the immediate option or the pre-commitment option on the response card, and zero otherwise. Here, we obtained qualitatively similar results as the results of our primary analysis reported in the paper based on the administrative data. Specifically, using the mailing response data, we estimated that employees' probability of immediately increasing their contributions to the targeted plan was 1.6 percentage points lower in the Simultaneous Pre-commitment condition than in the No Pre-commitment condition (p = .075). This represents a 32.7% decrease in immediate adoption relative to the 4.9 percent of employees who immediately enrolled in the targeted plan in the No Pre-commitment condition. And, there was virtually no difference across conditions in employees' probability of increasing their contributions to the targeted savings plan (p = .390). The weaker statistical significance for the regression predicting immediate adoption is likely because this analysis relies on a smaller sample of 2,029 employees and thus has less power than our primary analysis using the full sample across four universities.

Table A1-8. Measuring Decisions with Cards Mailed back to the One University that Tracked Responses

	Model 1: Immediate Adoption	Model 2: Overall Adoption
Simultaneous Pre-commitment	-.016*	-.008
	(.009)	(.009)
R-squared	.027	.024
Observations	2,029	2,029

Standard errors robust to heteroskedasticity in parentheses. The regressions control for gender, age decile, tenure decile, salary decile, faculty status, and birth month. We allowed the coefficients on the control variables to vary by university.
$*p \leq .10$, $**p \leq .05$, $***\ p \leq .01$.

Measuring overall adoption differently (Table A1-9). We calculated overall adoption using an alternative method that was meant to capture employees' direct responses to the mailers. For this measure, we only counted someone as enrolling if the first time their contribution rate increased (relative to their rate in September 2013) corresponded with one of the month(s) offered to them in the mailing. Specifically, for employees in the No Pre-commitment condition, this alternative measure of overall adoption equaled one if their first contribution rate increase was in November 2013 (corresponding to the immediate option), and zero otherwise. For employees in the Simultaneous Pre-commitment condition, overall adoption equaled one if their first increase was in November 2013 or in the pre-commitment month offered in their mailing (e.g., February 2014 for employees whose pre-commitment option was "in four months"), and zero otherwise. Of note, this method only modifies how we measured overall adoption; immediate adoption was always measured based on whether an employee had a contribution rate increase in November 2013 (compared to September 2013). When we estimated our primary regression specification with this alternate measure of overall adoption as the dependent variable, we obtained substantively similar results.

Table A1-9. Measuring Overall Adoption Differently

	Model 1: Overall Adoption
Simultaneous Pre-commitment	-.003
	(.007)
R-squared	.075
Observations	5,196

Standard errors robust to heteroskedasticity in parentheses. The regressions control for gender, age decile, tenure decile, salary decile, faculty status, and birth month. We allowed the coefficients on the control variables to vary by university.
*$p \leq .10$, **$p \leq .05$, *** $p \leq .01$.

Running logistic and fractional logit regressions (Table A1-10). We used logistic regressions rather than OLS to predict our binary dependent measures (immediate and overall adoption). We also used fractional logistic regressions rather than OLS to predict average savings rate during the study period—a continuous dependent measure ranging from 0% to 100%. Our results are not meaningfully changed.

Table A1-10. Running Logistic and Fractional Logit Regressions

	Model 1: Immediate Adoption	Model 2: Overall Adoption	Model 3: Average Savings Rate
Simultaneous Pre-commitment	-.359***	-.099	-.163**
	(.118)	(.092)	(.081)
Observations	5,050[a]	5,196	5,196

Standard errors robust to heteroskedasticity in parentheses. The regressions control for gender, age decile, tenure decile, salary decile, faculty status, and birth month. We allowed the coefficients on the control variables to vary by university.
*$p \leq .10$, **$p \leq .05$, *** $p \leq .01$.

[a] Some observations are automatically dropped out of the logistic regression that predicts immediate adoption (Model 1) because employees in one university x age decile group and one university x tenure decile group all had the value of zero for immediate adoption, leading controls for these groups to perfectly predict immediate adoption (and thus to be inestimable).

Dropping those with missing data (Table A1-11). For the results reported in the paper, we assigned a value of zero to an employee's contribution rate in a pay cycle if there was missing data for salary or contributions for that pay cycle. We do so because missing values for salary or contributions are likely to reflect short-term leaves of absence. Our results are robust if we instead drop employees who have missing data for salary or contributions in all months from November 2013 through June 2014.

Table A1-11. Dropping Those with Missing Data

	Model 1: Immediate Adoption	Model 2: Overall Adoption	Model 3: Average Savings Rates
Simultaneous Pre-commitment	-.020***	-.010	-.0015**
	(.007)	(.009)	(.001)
R-squared	.07	.08	.12
Observations	4,938	4,938	4,938

Standard errors robust to heteroskedasticity in parentheses. The regressions control for gender, age decile, tenure decile, salary decile, faculty status, and birth month. We allowed the coefficients on the control variables to vary by university.
*$p \leq .10$, **$p \leq .05$, *** $p \leq .01$.

Dropping those with missing data in key months (Table A1-12). Our results are robust if we instead drop employees who have missing data for salary or contributions in one of the key months used to calculate the immediate and overall adoption variables. Specifically, in the regressions reported below, we drop observations with missing salary or contributions data in September 2013 or November 2013 in Model 1, and we drop observations with missing salary or contributions data in September 2013 or June 2014 in Model 2.

Table A1-12. Dropping Those with Missing Data in Key Months

	Model 1: Immediate Adoption	Model 2: Overall Adoption
Simultaneous Pre-commitment	-.021***	-.009
	(.008)	(.010)
R-squared	.07	.09
Observations	4,618	4,543

Standard errors robust to heteroskedasticity in parentheses. The regressions control for gender, age decile, tenure decile, salary decile, faculty status, and birth month. We allowed the coefficients on the control variables to vary by university. *$p \leq .10$, **$p \leq .05$, *** $p \leq .01$.

Varying the cutoff for a contribution rate increase (Table A1-13). To calculate an employee's contribution rate in a given month, we divided her dollar contributions by her salary in that month. Since we constructed the overall adoption variable by comparing these imputed contribution rates in June 2014 versus September 2013, we want to ensure that our results are not spuriously driven by how we rounded our imputed contribution rates. For example, if an employee had an imputed contribution rate of 5.030% in September 2013 and 5.033% in June 2014, it is unlikely that this employee increased her contribution rate by .003%; rather, this difference in imputed contribution rates likely reflects a rounding issue. The same issue applies to our construction of the immediate adoption variable. For our main analyses in the manuscript, we only counted a contribution rate increase to be real if the increase was ≥ 1%. As a robustness check, we counted a contribution rate increase to be real if the increase was ≥ .10% or ≥ .01%, and we obtained similar results.

Table A1-13. Varying the Cutoff for a Contribution Rate Increase

Threshold for defining a real increase	Increase in contribution rate ≥ .1%		Increase in contribution rate ≥ .01%	
	Model 1: Immediate Adoption	Model 2: Overall Adoption	Model 3: Immediate Adoption	Model 4: Overall Adoption
Simultaneous Pre-commitment	-.020***	-.009	-.021***	-.008
	(.007)	(.009)	(.007)	(.009)
R-squared	.08	.09	.08	.09
Observations	5,196	5,196	5,196	5,196

Standard errors robust to heteroskedasticity in parentheses. The regressions control for gender, age decile, tenure decile, salary decile, faculty status, and birth month. We allowed the coefficients on the control variables to vary by university.

*$p \leq .10$, **$p \leq .05$, *** $p \leq .01$.

Including limited controls (Table A1-14). When controlling for the interaction between university and birth month (because of our stratified random assignment) but no other individual characteristics (i.e., gender, age, tenure, salary, and faculty status), our results remain qualitatively unchanged in magnitude, though the statistical significance is weaker for regressions predicting average savings rate.

Table A1-14. Including Limited Controls

	Model 1: Immediate Adoption	Model 2: Overall Adoption	Model 3: Average Savings Rate
Simultaneous Pre-commitment	-.017**	-.007	-.0012*
	(.007)	(.009)	(.001)
R-squared	.03	.04	.07
Observations	5,196	5,196	5,196

Standard errors robust to heteroskedasticity in parentheses. The regressions control for birth month and we allowed the coefficients on the birth month controls to vary by university. *$p \leq .10$, **$p \leq .05$, *** $p \leq .01$.

Web Appendix F. Summary of Differences Between Study 1 and Beshears et al. (2021)

Table A1-15 summarizes how the research questions, conditions, and findings differ between Study 1 and a companion paper, Beshears et al. (2021).

Table A1-15. Comparison of Study 1 with Beshears et al. (2021)

	Study 1 in the Current Paper	Beshears et al. (2021)
Research Question	What is the effect of simultaneous pre-commitment on retirement savings decisions?	What is the effect of fresh-start framing on retirement savings decisions?
Conditions Included in Analysis	Simultaneous Pre-commitment VS. No Pre-commitment	Simultaneous Pre-commitment Condition Linked with Temporal Landmarks VS. Simultaneous Pre-commitment[a]
Findings	Simultaneous pre-commitment reduced immediate adoption, had a null effect on overall adoption, and reduced average savings rates.	Fresh-start framing had a null effect on immediate adoption, increased overall adoption, and increased average savings rates.
Theorized Mechanism	Simultaneous pre-commitment decreases inferred urgency	Fresh start framing heightens motivation to initiate new goals

[a] Beshears et al. (2021) draw a distinction between temporal landmarks that feel like the beginning of a new time period and are associated with fresh starts (Birthdays, New Year’s, the first day of spring) and temporal landmarks that are not associated with fresh starts (Thanksgiving, Martin Luther King Day, Valentine’s Day); the latter are used for placebo regressions in their paper.

Web Appendix G. Post-test Following Study 1

Methods

We recruited participants through Amazon's Mechanical Turk (MTurk) to take a short survey for a pre-registered study. The 1,499 participants (M_{age} = 36.41, SD_{age} = 11.59; 47.3% female) who successfully answered comprehension check questions were included in our study, as stipulated in our pre-registration. All participants were asked to imagine that the HR department at "Company X" planned to send its employees mailings about the company's retirement savings program. Participants were randomly assigned to read about and evaluate one of three mailings: a No Pre-commitment mailing, a Simultaneous Pre-commitment mailing, or a Sequential Pre-commitment mailing.

The No Pre-commitment mailing depicted a simplified version of the No Pre-commitment mailing from our field experiment, adapted such that all references to universities and specific retirement savings plans were replaced with references to Company X and its hypothetical savings program. The No Pre-commitment mailing encouraged employees to sign up for Company X's retirement savings plan immediately.

The Simultaneous Pre-commitment mailing depicted a simplified version of the Simultaneous Pre-commitment mailing from our field experiment, again adapted to reference only Company X and its hypothetical savings program. The Simultaneous Pre-commitment mailing offered employees two options: the option to start contributing to Company X's savings program immediately and the option to pre-commit to start contributing in four months.

The Sequential Pre-commitment mailing was comprised of two mailings sent in stages. In the first stage, the mailing depicted was identical to the No Pre-commitment mailing, which invited employees to enroll in Company X's savings program immediately. Participants learned that in the second stage, if an employee did not reply to the initial No Pre-commitment mailing, the HR department would send a follow-up mailing with an invitation to pre-commit to enroll in the savings program in four months. This second-stage mailing explained to employees that they received the second offer because they did not respond to the first-stage mailing, and it only presented the option for employees to pre-commit to enroll in four months.

After participants read the mailing(s) associated with their experimental condition, we assessed their understanding with a series of comprehension check questions, and we ended the survey for participants who failed these questions (N = 83). 95% of participants passed our comprehension check. The rate did not differ significantly between the No Pre-commitment condition (97%) and the Simultaneous Pre-commitment condition (96%; p = .60), but it was slightly lower in the Sequential Pre-commitment condition than the other conditions (93%; both p's < .05), likely because this comprehension check involved more questions than the others.

Next, we measured *inferred urgency* by asking: "To what extent will this mailing convey to employees that the human resources staff urgently recommends that employees enroll in the retirement savings program?" (1 = Not at all urgently, 7 = Very urgently). We also included a measure about inferred future opportunities to save based on an earlier version of our theory. We asked: "To what extent will this mailing suggest to employees that they will have other, future opportunities to enroll in the retirement savings program?" (1 = No future opportunities suggested, 7 = Future opportunities strongly suggested).

Finally, we asked participants for their age and gender.

Results

Supporting Hypothesis 1, participants reported that the No Pre-commitment mailing conveyed a more urgent recommendation to save (M = 5.10, SD = 1.42) than the Simultaneous Pre-commitment mailing (M = 4.77, SD = 1.58, *t*(1,011) = 3.54, *p* < .001). This difference between the two mailings may help to explain why our Simultaneous Pre-commitment mailing failed to increase savings and also reduced immediate adoption rates compared to the No Pre-commitment mailing in our field experiment.

Also consistent with Hypothesis 1, participants reported that the Sequential Pre-commitment mailing conveyed a marginally more urgent recommendation to save (M = 5.27, SD = 1.30) than the No Pre-commitment mailing (M = 5.10, SD = 1.42; *t*(993) = 1.89; *p* = .06). Also, as expected, participants reported that the Sequential Pre-commitment mailing conveyed a more urgent recommendation than the Simultaneous Pre-commitment mailing (M = 4.77, SD = 1.58; *t*(988) = 5.40, *p* < .001).

Based on a previous version of our theory, we hypothesized in our pre-registration and confirmed that, compared to the No Pre-commitment condition (M = 3.09, SD = 1.73), people inferred that there would be more future opportunities to save in both the Simultaneous Pre-commitment (M = 3.96, SD = 2.00), *t*(1,011) = 7.48, *p* < .001) and Sequential Pre-commitment conditions (M = 3.92, SD = 1.92, t(993) = 7.19, *p* < .001).

Web Appendix H. Exploratory Factor Analysis (Study 2)

We conducted an exploratory factor analysis on six items: two urgency items and four decision difficulty items. Below are the six items (with shorthand labels in bold). All items use a 7-point Likert scale ranging from 1 = Not at all to 7 = Very much.

Inferred urgency item 1: To what extent do you think Company X recommends that employees enroll in the benefits programs as soon as they can?
Inferred urgency item 2: To what extent do you think Company X urgently recommends that employees enroll in the benefits programs?
Decision difficulty item 1: To what extent were you overwhelmed when you were making your enrollment decision?
Decision difficulty item 2: To what extent were you frustrated when you were making your enrollment decision?
Decision difficulty item 3: To what extent were you annoyed when you were making your enrollment decision?
Decision difficulty item 4: To what extent did you find the choice difficult when you were making your enrollment decision?

To conduct the exploratory factor analysis, we used Ordinary Least Squares (OLS) with an oblimin rotation to find the minimum residual solution. See the factor loadings presented in Table A1-16 below. Note, factor loadings below .5 are not displayed for ease of visualization.

Table A1-16. Factor Loadings

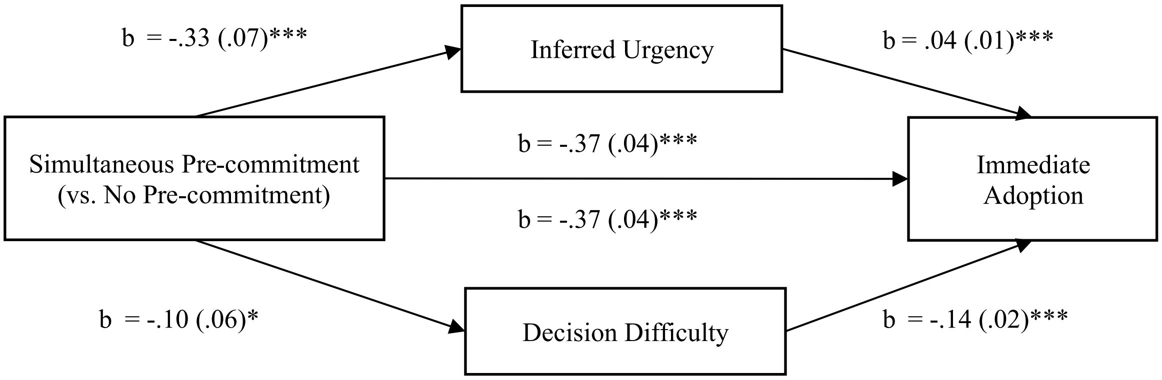
	Factor1	Factor2
Inferred urgency item 1		.77
Inferred urgency item 2		.91
Decision difficulty item 1	.63	
Decision difficulty item 2	.91	
Decision difficulty item 3	.83	
Decision difficulty item 4	.72	

As shown in Table A1-16, the items used to measure "inferred urgency" and the items to measure "decision difficulty" loaded on separate factors. Also, the "inferred urgency" composite (average of the two items) and the "decision difficulty" composite (average of the four items) have a relatively low correlation of Pearson $r = .15$. Together these observation suggests that the two constructs are distinct.

Web Appendix I. Multi-Mediator Models (Study 2)

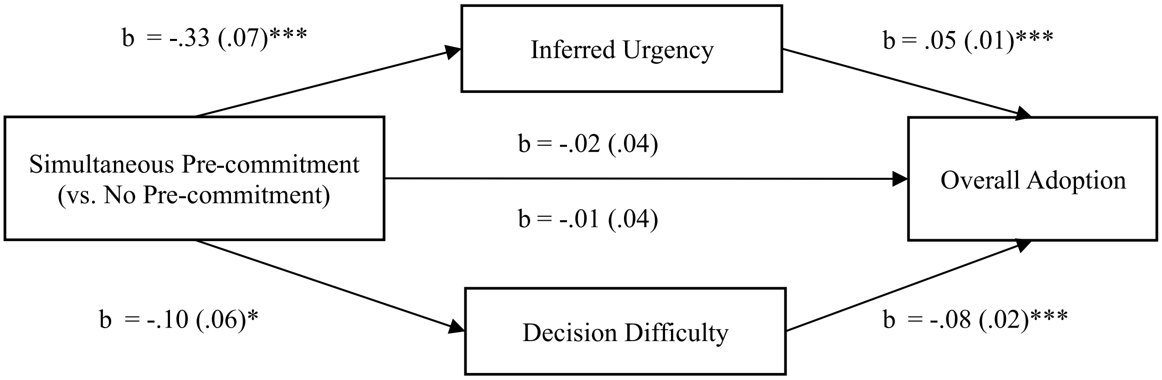
Figures A1-2-A1-5 depict how the hypothesized mediators explain the effects of offering simultaneous pre-commitment and sequential pre-commitment on immediate and overall adoption of the employee benefits programs. All regression coefficients are unstandardized, and standard errors are presented in parentheses. The coefficients above the paths from Simultaneous Pre-commitment and Sequential Pre-commitment to "immediate adoption" and "overall adoption" represent the total effects and the coefficients below the paths represents the direct effects. Coefficients significantly different from zero are indicated by asterisks (*$p \leq .10$, **$p \leq .05$, *** $p \leq .01$).

Figure A1-2. Simultaneous vs. No Pre-commitment: Immediate Adoption Mediation



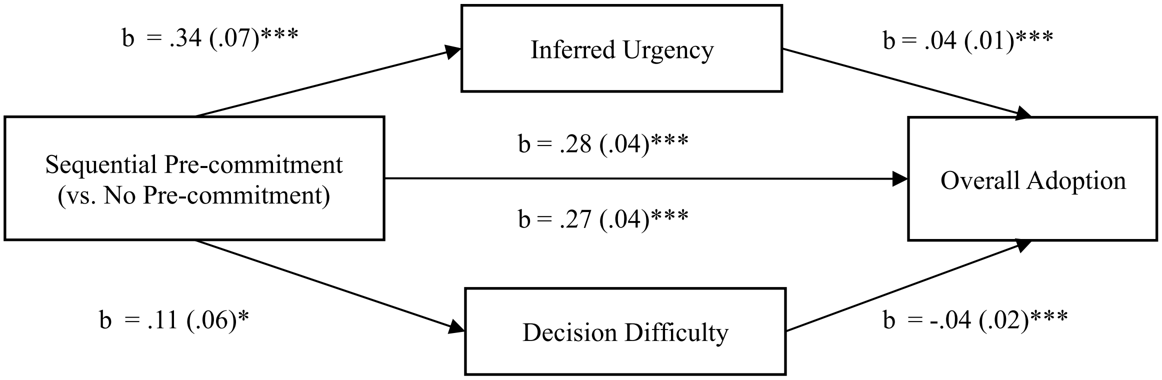
Indirect effects	*b*	*95% CI*
Inferred Urgency	-.01	[-.02, -.005]
Decision Difficulty	.01	[-.003, .03]

Figure A1-3. Simultaneous vs. No Pre-commitment: Overall Adoption Mediation



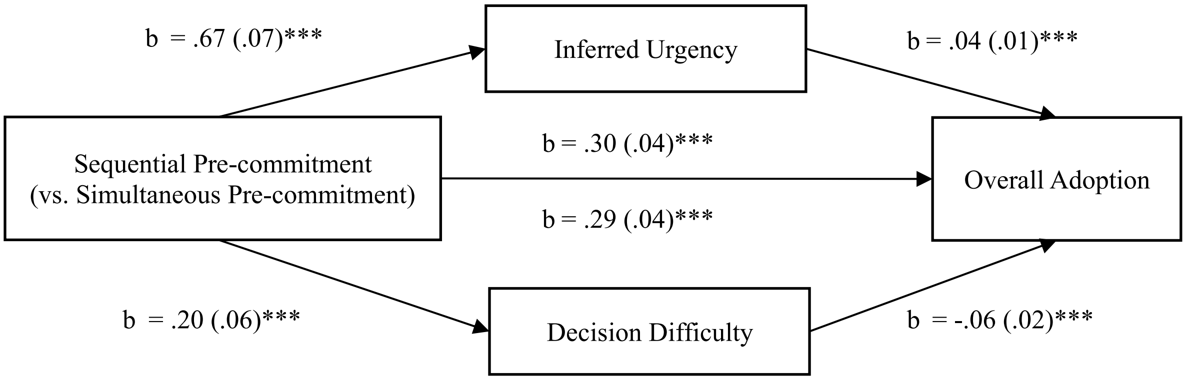
Indirect effects	*b*	*95% CI*
Inferred Urgency	-.02	[-.03, -.01]
Decision Difficulty	.01	[-.001, .02]

Figure A1-4. Sequential Pre-commitment vs. No Pre-commitment: Overall Adoption Mediation



Indirect effects	*b*	*95% CI*
Inferred Urgency	.01	[.004, .03]
Decision Difficulty	-.004	[-.01, .0003]

Figure A1-5. Sequential Pre-commitment vs. Simultaneous Pre-commitment: Overall Adoption Mediation



Indirect effects	*b*	*95% CI*
Inferred Urgency	.02	[.01, .04]
Decision Difficulty	-.01	[-.02, -.005]

Web Appendix J. Two-Condition Replication of Study 2

We conducted a laboratory experiment in which participants decided (hypothetically) whether and when to enroll in three different benefits programs offered by a new employer. This is an exact replication of Study 2 reported in the manuscript except it only included the No Pre-commitment and Simultaneous Pre-commitment conditions, and we used a different set of questions to measure mechanisms.

Methods

As stipulated in our pre-registration, we recruited MTurk participants who passed a brief attention check and were fully employed at a firm other than MTurk. A total of 1,161 participants satisfied these selection criteria and completed our pre-registered study (M_{age} = 38.24; SD_{age} = 10.61; 52.3% female).

The vignette about enrollment in Company X's HR benefits programs is identical to Study 2, however, this study only included the No Pre-commitment and Simultaneous Pre-commitment conditions.

After participants made their enrollment decisions, participants in both conditions answered two questions about Company X: "To what extent do you think Company X urgently recommends that employees enroll in the benefits programs" and "To what extent do you think Company X recommends that employees enroll in the benefits programs as soon as possible?" (1 = Not at all, 7 = Very much). These questions were collapsed into a single measure of *inferred urgency* ($r = .77, p < .001$). We next asked two questions measuring participants' *anticipated negative emotion* associated with enrolling in the benefits programs: "If you were a full-time employee at Company X, how painful would it be for you to enroll in any of Company X's benefits programs?" (1 = Not at all painful, 7 = Very painful) and "If you were a full-time employee at Company X, how unpleasant would it be for you to enroll in any of Company X's benefits programs?" (1 = Not at all unpleasant, 7 = Very unpleasant). These two items were collapsed into a single measure ($r = .74, p < .001$).

We also asked three questions designed to assess potential alternative mechanisms for our predicted results. Specifically, we measured *decision difficulty* with one item: "How difficult was it for you to make this enrollment decision?" (1 = Not at all difficult, 7 = Very difficult). We measured *confusion* with one item: "How confusing was Company X's enrollment process?" (1 = Not at all confusing, 7 = Very confusing). And we measured the *inferred thoughtfulness* of Company X's enrollment process with one item: "How much thought do you think Company X put into the design of the enrollment process?" (1 = Very little thought, 7 = A lot of thought). At the end of our study, participants were asked about their age, gender, education, and income.

Results

Inferred Urgency. Replicating the results of Study 2, compared to those in the No Pre-commitment condition (M = 4.15, SD = 1.70), participants in the Simultaneous Pre-commitment condition rated Company X's implied recommendation to enroll as less urgent (M = 3.56, SD = 1.73; $b = -.59, p < .001$).

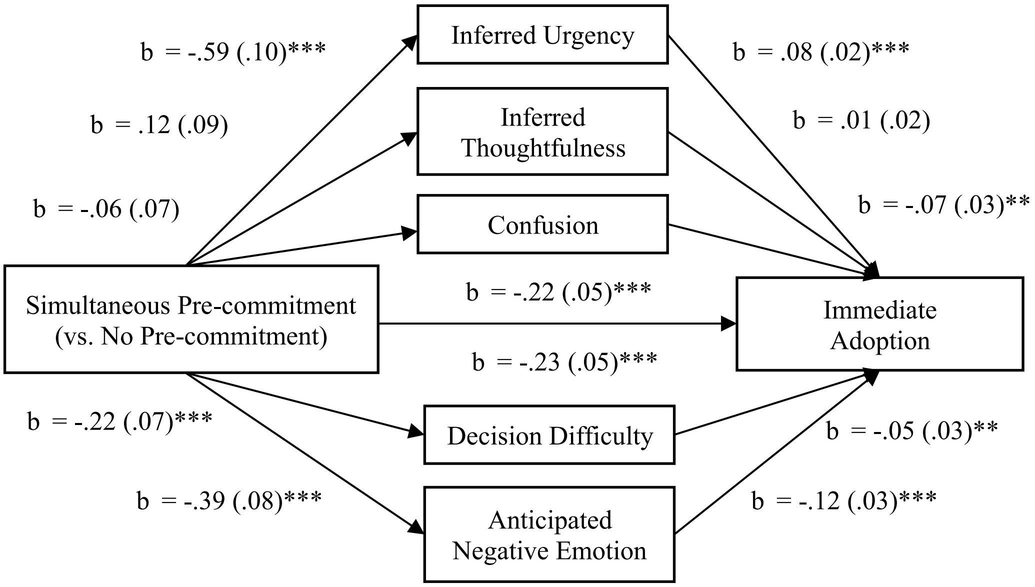
Immediate Adoption. Replicating the results from our field experiment and Study 2, participants in the Simultaneous Pre-commitment condition signed up for fewer benefits immediately (M = 1.77, SD = .96) than participants in the No Pre-commitment condition (M = 1.99, SD = .86; b = -.22, $p < .001$). We estimated a multi-mediator model including all the other potential mediators listed in the methods section, and inferred urgency mediated the effect of simultaneous pre-commitment (vs. no pre-commitment) on immediate adoption (b = -.05; 95% CI = [-.07, -.03]; See Figure A1-6).

Overall Adoption. Consistent with the results of our field experiment and Study 2, compared to the No Pre-commitment condition (M = 1.99, SD = .86), offering simultaneous pre-commitment (M = 2.03, SD = .89) did not significantly change the total number of benefits programs people enrolled in (b = .04, *p* = .401). And in a multi-mediator model involving other potential mediators, inferred urgency mediated the effect of simultaneous pre-commitment (vs. no pre-commitment) on overall adoption (b = -.06, 95% CI = [-.08, -.03]; See Figure A1-7). Also, consistent with previous research that has theorized about the positive impact of pre-commitment, we found evidence that anticipated negative emotions (i.e., painfulness, unpleasantness) *positively* mediated the effect of simultaneous pre-commitment on overall adoption (b = .04, 95% CI = [.02, .07]). That is, when people had the option to enroll in the benefits at a delay (vs. when they could only enroll immediately), they perceived that enrolling would feel less aversive, which predicted greater overall enrollment. Thus, consistent with our theory, the null effect of simultaneous pre-commitment on overall adoption observed in this study may be explained by competing mechanisms: the negative indirect effect via inferred urgency may have negated the positive indirect effect via anticipated negative emotions.

Web Appendix K. Multi-Mediator Models (Two-Condition Replication of Study 2)

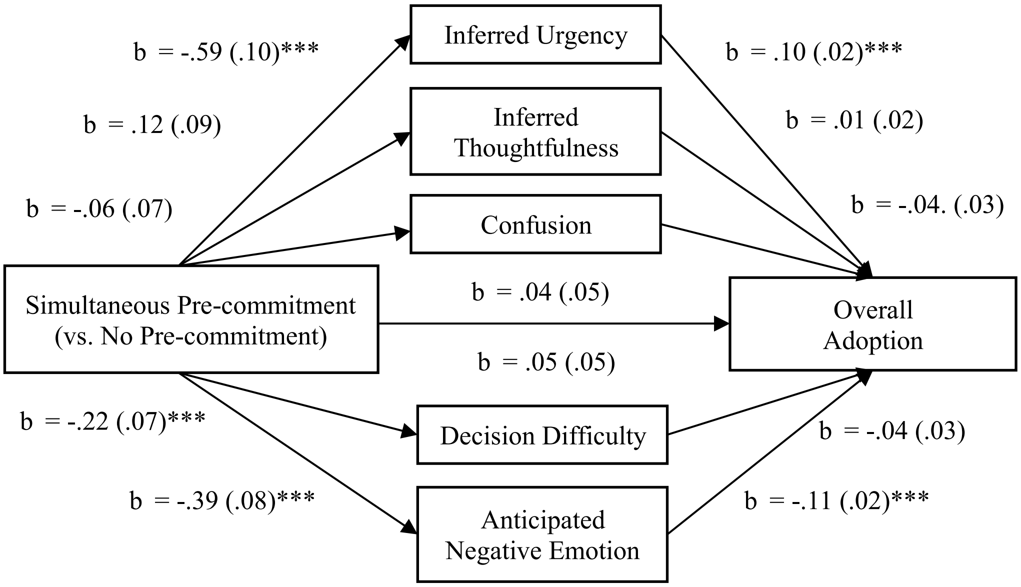
Figures A1-6 and A1-7 depict how the hypothesized mediators explain the effects of offering simultaneous pre-commitment on immediate and overall adoption of the employee benefits programs. All regression coefficients are unstandardized, and standard errors are presented in parentheses. The coefficients above the paths from Simultaneous Pre-commitment to "immediate adoption" and "overall adoption" represent the total effects and the coefficients below the paths represents the direct effects. Coefficients significantly different from zero are indicated by asterisks (*$p \leq .10$, **$p \leq .05$, *** $p \leq .01$).

Figure A1-6. Simultaneous vs. No Pre-commitment: Immediate Adoption Mediation



Indirect effects	*b*	*95% CI*
Inferred urgency	-.05	[-.07, -.03]
Anticipated negative emotion	.05	[.02, .07]
Inferred thoughtfulness	.0001	[-.004, .007]
Confusion	.003	[-.005, .02]
Decision difficulty	.01	[.0001, .03]

Figure A1-7. Simultaneous vs. No Pre-commitment: Overall Adoption Mediation

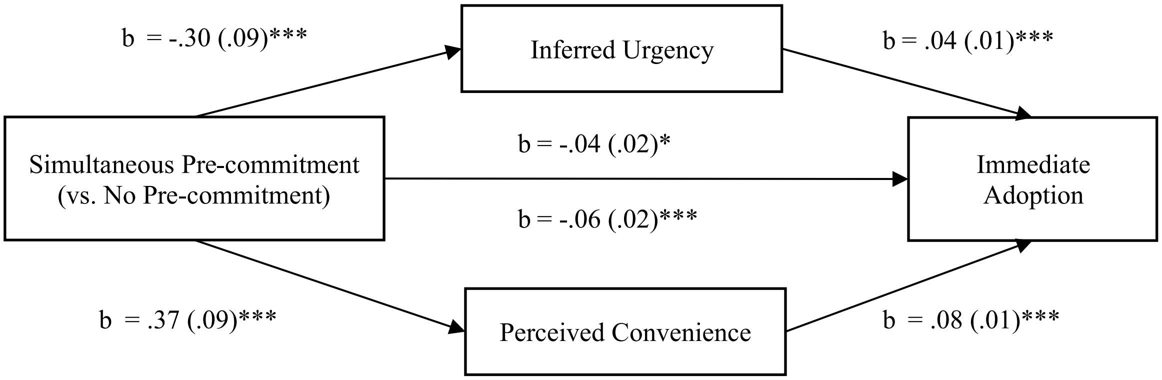


Indirect effects	*b*	*95% CI*
Inferred urgency	-.06	[-.08, -.03]
Anticipated negative emotion	.04	[.02, .07]
Inferred thoughtfulness	.0004	[-.005, .007]
Confusion	.002	[-.003, .01]
Decision difficulty	.01	[-.002, .02]

Web Appendix L. Multi-Mediator Models (Study 3)

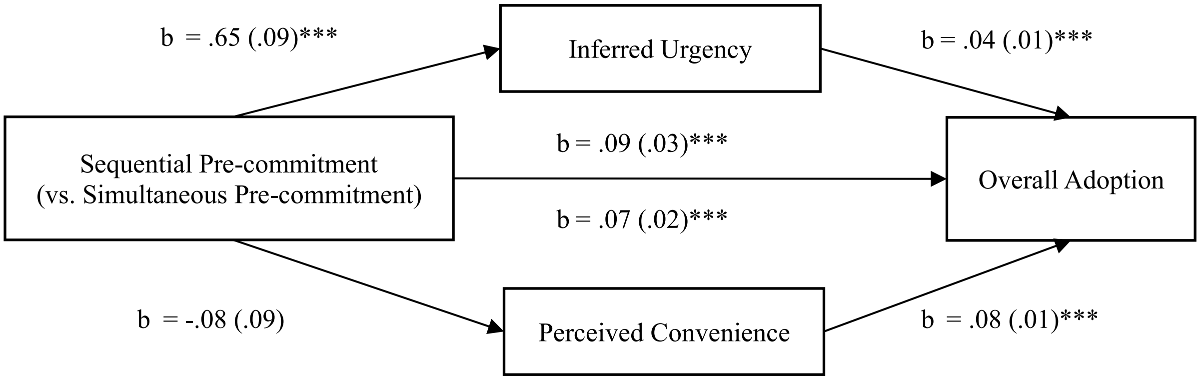
Figures A1-8 and A1-9 depict how the hypothesized mediators explain the effects of offering simultaneous pre-commitment and sequential pre-commitment on immediate and overall adoption of the financial well-being assessment. All regression coefficients are unstandardized, and standard errors are presented in parentheses. The coefficients above the paths from Simultaneous Pre-commitment and Sequential Pre-commitment to "immediate adoption" and "overall adoption" represent the total effects and the coefficients below the paths represents the direct effects. Coefficients significantly different from zero are indicated by asterisks (*$p \leq .10$, **$p \leq .05$, *** $p \leq .01$).

Figure A1-8. Simultaneous vs. No Pre-commitment: Immediate Adoption Mediation



Indirect effects	*b*	*95% CI*
Inferred Urgency	-.01	[-.02, -.004]
Perceived Convenience	.03	[.02, .05]

Figure A1-9. Sequential Pre-commitment vs. Simultaneous Pre-commitment: Overall Adoption Mediation



Indirect effects	*b*	*95% CI*
Inferred Urgency	.03	[.02, .04]
Perceived Convenience	-.01	[-.02, .01]

REFERENCES

Ashraf, Nava, Dean Karlan, and Wesley Yin (2006), “Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines,” *The Quarterly Journal of Economics*, 121 (2), 635–672.

Benartzi, Shlomo (2001), “Excessive Extrapolation and the Allocation of 401 (k) Accounts to Company Stock,” *The Journal of Finance*, 56 (5), 1747-1764.

Benartzi, Shlomo, and Richard H. Thaler (2013), “Behavioral Economics and the Retirement Savings Crisis,” *Science,* 339 (6124), 1152-1153.

Beshears, John, Hengchen Dai, Katherine L. Milkman, and Shlomo Benartzi (2021), “Using Fresh Starts to Nudge Increased Retirement Savings,” *Organizational Behavior and Human Decision Processes*, 167, 72-87.

Brown, Christina L., and Aradhna Krishna (2004), “The Skeptical Shopper: A Metacognitive Account for the Effects of Default Options on Choice,” *Journal of Consumer Research*, 31 (3), 529–539.

Choi, James J. (2015), “Contributions to Defined Contribution Pension Plans,” *Annual Review of Financial Economics*, 7 (1), 161–178.

D'Adda, Giovanna, Arianna Galliera, and Massimo Tavoni (2020), “Urgency and Engagement: Empirical Evidence from a Large-Scale Intervention on Energy Use Awareness,” *Journal of Economic Psychology*, 102275.

Fox, Craig R., Rebecca K. Ratner, and Daniel S. Lieb (2005), "How Subjective Grouping of Options Influences Choice and Allocation: Diversification Bias and the Phenomenon of Partition Dependence," *Journal of Experimental Psychology: General,* 134 (4), 538.

Frederick, Shane, George Loewenstein, and Ted O'Donoghue (2002), “Time Discounting and

Time Preference: A Critical Review," *Journal of Economic Literature*, 40 (2), 351-401.

Goodman, Joseph K., Susan M. Broniarczyk, Jill G. Griffin, and Leigh McAlister (2013), "Help or Hinder? When Recommendation Signage Expands Consideration Sets and Heightens Decision Difficulty," *Journal of Consumer Psychology*, 23 (2), 165-174.

Kardes, Frank R., Steven S. Posavac, and Maria L. Cronley (2004), "Consumer Inference: A Review of Processes, Bases, and Judgment Contexts," *Journal of Consumer Psychology*, 14 (3), 230–256.

Karlan, Dean, and Leigh L. Linden (2014), "Loose Knots: Strong Versus Weak Commitments to Save for Education in Uganda," NBER Working Paper.

Knoll, Melissa A. Z., and Carrie R. Houts (2012), "The Financial Knowledge Scale: An Application of Item Response Theory to the Assessment of Financial Literacy," *Journal of Consumer Affairs*, 46 (3), 381–410.

Krijnen, Job M. T., David Tannenbaum, and Craig R. Fox (2017), "Choice Architecture 2.0: Behavioral Policy as an Implicit Social Interaction," *Behavioral Science & Policy*, 3 (2), i–18.

Lieberman, Alicea, Kristen E. Duke, and On Amir (2019), "How Incentive Framing Can Harness the Power of Social Norms," *Organizational Behavior and Human Decision Processes*, 151, 118–131.

McKenzie, Craig R. M., Michael J. Liersch, and Stacey R. Finkelstein (2006), "Recommendations Implicit in Policy Defaults," *Psychological Science*, 17 (5), 414–420.

Milkman, Katherine L., Todd Rogers, and Max H. Bazerman (2008), "Harnessing Our Inner Angels and Demons: What We Have Learned About Want/Should Conflicts and How That Knowledge Can Help Us Reduce Short-Sighted Decision Making," *Perspectives on*

Psychological Science, 3 (4), 324–338.

Milkman, Katherine L., Todd Rogers, and Max H. Bazerman (2009), “Highbrow Films Gather Dust: Time-Inconsistent Preferences and Online DVD Rentals,” *Management Science*, 55 (6), 1047–1059.

Milkman, Katherine L., Todd Rogers, and Max H. Bazerman (2010), “I’ll Have the Ice Cream Soon and the Vegetables Later: A Study of Online Grocery Purchases and Order Lead Time,” *Marketing Letters*, 21 (1), 17–35.

Morewedge, Carey K., Leif Holtzman, and Nicholas Epley (2007), “Unfixed Resources: Perceived Costs, Consumption, and the Accessible Account Effect,” *Journal of Consumer Research*, 34 (4), 459-467.

O'Donoghue, Ted, and Matthew Rabin (1999), “Doing It Now or Later,” *American Economic Review*, 89 (1), 103–124.

Read, Daniel, and Barbara Van Leeuwen (1998), “Predicting Hunger: The Effects of Appetite and Delay on Choice,” *Organizational Behavior and Human Decision Processes*, 76 (2), 189–205.

Read, Daniel, George Loewenstein, and Shobana Kalyanaraman (1999), “Mixing Virtue and Vice: Combining the Immediacy Effect and the Diversification Heuristic,” *Journal of Behavioral Decision Making*, 12 (4), 257–273.

Reiff, Joseph S., Justin Zhang, Jana Gallus, Hengchen Dai, Nathaniel Pedley, Sitaram Vangala, Richard Leuchter, Gregory Goshgarian, Craig R. Fox, Maria Han, & Daniel Croymans (2022). “When peer comparison information harms physician well-being,” *Proceedings of the National Academy of Sciences of the United States of America,* 119 (29), e2121730119.

Rogers, Todd, and Max H. Bazerman (2008), “Future Lock-In: Future Implementation Increases Selection of ‘Should’ Choices,” *Organizational Behavior and Human Decision Processes*, 106 (1), 1–20.

Rogers, Todd, Katherine Milkman, and Kevin Volpp (2014), “Commitment devices: Using Initiatives to Change Behavior,” *Journal of the American Medical Association*, 311 (20), 2065-2066.

Schwartz, Janet, Daniel Mochon, Lauren Wyper, Josiase Maroba, Deepak Patel, and Dan Ariely (2014), "Healthier by Precommitment," *Psychological Science*, 25 (2), 538-546.

Smith, N. Craig, Daniel G. Goldstein, and Eric J. Johnson (2013), “Choice Without Awareness: Ethical and Policy Implications of Defaults,” *Journal of Public Policy & Marketing*, 32 (2), 159–172.

Tannenbaum, David, Craig R. Fox, and Noah J. Goldstein (2013), "Partitioning Menu Items to Nudge Single-Item Choice," working paper.

Thaler, Richard H., and Shlomo Benartzi (2004), “Save More Tomorrow™: Using Behavioral Economics to Increase Employee Saving,” *Journal of Political Economy*, 112 (S1), S164–S187.

Ubel, Peter A., David A. Comerford, and Eric Johnson (2015), “Healthcare.gov 3.0—Behavioral Economics and Insurance Exchanges,” *New England Journal of Medicine*, 372 (8), 695–698.

Valenzuela, Ana, and Priya Raghubir (2009), “Position-Based Beliefs: The Center-Stage Effect,” *Journal of Consumer Psychology*, 19 (2), 185–196.

VanEpps, Eric M., Julie S. Downs, and George Loewenstein (2016), “Advance Ordering for Healthier Eating? Field Experiments on the Relationship Between the Meal Order–

Consumption Time Delay and Meal Content," *Journal of Marketing Research*, 53 (3), 369–380.

Zauberman, Gal, and John G. Lynch Jr. (2005), "Resource Slack and Propensity to Discount Delayed Investments of Time versus Money," *Journal of Experimental Psychology: General*, 134 (1), 23-37.

Zhu, Meng, Yang Yang, and Christopher K. Hsee (2018), "The Mere Urgency Effect," *Journal of Consumer Research*, 45 (3), 673–690.

CHAPTER 2: WHEN PEER COMPARISON INFORMATION HARMS WELL-BEING

Joseph S. Reiff (co-first author) [a]

Justin C. Zhang (co-first author) [b]

Jana Gallus [a]

Hengchen Dai [a]

Nathaniel M. Pedley [c]

Sitaram Vangala [d]

Richard K. Leuchter [d]

Gregory Goshgarian [e]

Craig R. Fox [a]

Maria Han [d]

Daniel M. Croymans [c,e]

[a] University of California, Los Angeles
[b] University of California, San Francisco
[c] UCLA Health
[d] UCLA Health Department of Medicine
[e] University of Maryland, College Park
[e] SCAN Health

Acknowledgments: We thank UCLA Health Department of Medicine for allowing us to conduct this intervention and assisting with data collection. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1650604. We thank Jose Cervantez for his outstanding research assistance.

Abstract: Policymakers and business leaders often use peer comparison information—showing people how their behavior compares to that of their peers—to motivate a range of behaviors. Despite their widespread use, the potential impact of peer comparison interventions on recipients' well-being is largely unknown. We conducted a five-month field experiment involving 199 primary care physicians and 46,631 patients to examine the impact of a peer comparison intervention on physicians' job performance, job satisfaction, and burnout. We varied whether physicians received information about their preventive care performance compared to that of other physicians in the same health system. Our analyses reveal that our implementation of peer comparison did not significantly improve physicians' preventive care performance, but it did significantly decrease job satisfaction and increase burnout, with the effect on job satisfaction persisting for at least four months after the intervention had been discontinued. Quantitative and qualitative evidence on the mechanisms underlying these unanticipated negative effects suggest that the intervention inadvertently signaled a lack of support from leadership. Consistent with this account, providing leaders with training on how to support physicians mitigated the negative effects on well-being. Our research uncovers a critical potential downside of peer comparison interventions, highlights the importance of evaluating the psychological costs of behavioral interventions, and points to how a complementary intervention—leadership support training—can mitigate these costs.

INTRODUCTION AND THEORY

Many behavior change interventions leverage peer comparison information, which involves showing people how their behavior compares to that of their peers. Peer comparison interventions have successfully improved educational outcomes (Tran and Zeckhauser 2012), reduced energy consumption (Allcott and Rogers 2014), boosted voter turnout (Gerber and Rogers 2009), increased charitable giving (Frey and Meier 2004), and bolstered employee productivity (Blader, Gartenberg, and Prat 2020). Within healthcare systems, peer comparison interventions targeting physicians have curbed overprescribing of antibiotics (Meeker et al. 2016), improved emergency department efficiency (Song et al. 2018), and increased adherence to best practices (Navathe et al. 2020). Previous research has primarily focused on how peer comparison interventions affect targeted behaviors. Yet, by only focusing on these behaviors, researchers and policymakers risk overlooking an important, less-visible class of outcomes: recipients' well-being.

The original goal of the current research was to evaluate whether a newly introduced peer comparison intervention would improve physicians' preventive care performance. In a natural field experiment within a large hospital system, we found no evidence of such an effect on physician performance. However, we observed an unexpected negative impact of the peer comparison intervention on physicians' job satisfaction and burnout. The primary goal of this paper is to understand these harmful effects so that they can be avoided in the future.

Recent research suggests that peer comparison information can be aversive to recipients (Allcott and Kessler 2019). In particular, being compared to higher ranked peers can be discouraging (Brown et al. 2007; Lockwood and Kunda 1997; Rogers and Feller 2016), resulting in feelings of shame (Butera et al. 2022) or stress (Hermes et al. 2021). Extending prior work that has focused on immediate affective reactions to upward social comparisons, we theorize that, when implemented in organizational contexts, peer comparison interventions can elicit another psychological process and impose long-term psychological costs. We propose that the use of peer comparison interventions can alter workers' perceptions of and relationships with the leaders implementing the intervention as they try to make sense of how and why this information is being presented to them (Krijnen, Tannenbaum, and Fox 2017). Workers may perceive their leaders' use of the intervention as reflecting inadequate leadership support if workers deem that the intervention's design and implementation violate existing norms of cooperation (Blader, Gartenberg, and Prat 2020; Gallus et al. 2021) or contradict workers' beliefs about what constitutes appropriate performance feedback. Given that leadership support is key to work-related well-being[1] (Bobbio, Bellan, and Manganelli 2012; Maslach, Schaufeli, and Leiter 2001; Shanafelt et al. 2015; West, Dyrbye, and Shanafelt 2018), job satisfaction and burnout may be harmed by the use of peer comparison interventions.

These dynamics are particularly important to examine within the healthcare context, where public health leaders must balance dual objectives. As health insurance plans place greater weight on optimizing "healthcare quality" metrics, health systems across the US are increasingly tracking physician behavior and implementing behavioral interventions (e.g., using peer comparison information) in an attempt to improve performance on these metrics (Mayer, Venkatesh, and Berwick 2021; McKethan and Jha 2014). Even Medicare has administered large-scale programs that use peer comparison information ("Care Compare: Doctors and Clinicians Initiative" 2021; Hassol et al. 2021). Concurrently, almost half of physicians in the US report

---

[1] We use the phrase "work-related well-being," or "well-being" for short, to refer to employees' job satisfaction and burnout.

experiencing burnout (Kane 2021), which is associated with greater turnover, reduced job performance, increased alcohol abuse, and higher rates of suicide (George and Jones 1996; Judge et al. 2001; Staw, Sutton, and Pelled 1994; West, Dyrbye, and Shanafelt 2018; Wright and Bonett 1997)–estimated to cost the US healthcare system $5 billion annually (Han et al. 2019; Yates 2020).

FIELD EXPERIMENT TESTING PEER COMPARISON INTERVENTIONS

We conducted a 5-month field experiment (from November 2019 through March 2020) in partnership with UCLA Health to examine the impact of a peer comparison intervention on both physicians' job performance and well-being. The experiment involved 199 primary care physicians (PCPs) and their 46,631 patients. PCPs were cluster-randomized at the clinic level to one of three study conditions: Control (Condition 1), Peer Comparison (Condition 2), or Peer Comparison and Leadership Training (Condition 3). PCPs in all conditions received monthly emails from UCLA Health's department leadership with feedback about their preventive care performance. Their performance was summarized with a "health maintenance (HM) completion rate," which reflects the proportion of recommended preventive care measures, such as routine screenings, that were completed by their patients in the previous 3 months. The emails in the Control Condition only contained feedback about the PCP's personal score. The emails in the Peer Comparison Condition also contained a list of the month's "Top 25 Primary Care Physicians" as well as information about where the PCP fell in the performance distribution. PCPs in the Peer Comparison and Leadership Training Condition received the same emails as those used in the Peer Comparison Condition, but leaders at each clinic also participated in training on how to support their physicians' preventive care performance. See Materials and Methods and the *SI Appendix* for more information.

Order Rate of Preventive Screening Exams. For each patient who visited a PCP in our experiment, we tracked the share of recommended preventive measures that were ordered by their PCP within the seven days following the visit. This is our primary pre-registered outcome. The average order rates were 9.4% in the Control Condition (SD = 25.4%), 10.5% in the Peer Comparison Condition (SD = 26.5%), and 9.9% in the Peer Comparison and Leadership Training Condition (SD = 26.0%). Following our pre-registered analysis plan, we first compared PCPs' order rates between the Control condition and the conditions containing peer comparison information (Conditions 2 and 3) and found no statistically significant difference ($P = 0.143$). As an exploratory analysis, we also compared the order rates between Condition 1 and Condition 2 but still did not find any statistically significant differences ($P = 0.324$). The regression tables for these analyses are reported in Section 8 of the *SI Appendix*.

Previous research suggests that the impact of peer comparison may depend on baseline performance (Ashraf, Bandiera, and Lee 2014; Bandiera, Barankay, and Rasul 2013; Bogard et al. 2020), discouraging low performers while encouraging high performers. However, our post-hoc analysis found no evidence that the estimated effect of the peer comparison intervention on order rates was moderated by PCPs' baseline performance (i.e., the HM completion rate displayed in the first intervention email). See Section 9 of the *SI Appendix* for details.

Job Satisfaction and Burnout. Next, we examined differences between conditions in our two well-being outcomes, job satisfaction and burnout, which were measured by UCLA Health in quarterly surveys. We first confirmed that job satisfaction and burnout were balanced across conditions in the baseline period before the experiment started (October 2019; F-test for joint significance: job satisfaction, $P = 0.432$; burnout, $P = 0.134$). We then evaluated the effects of our interventions on job satisfaction and burnout at the end of the 5-month experimental period

(April 2020). The regression-estimated treatment effects are displayed in Figures 2-1 and 2-2. Since both the peer comparison and the leadership support training interventions could separately impact well-being, we first evaluated the effects of peer comparison alone (comparing Condition 2 with Condition 1). We then tested the effects of adding leadership training (Condition 3 vs. Condition 2). Compared to the Control Condition (job satisfaction M = 5.47, SD = 0.91; burnout M = 1.93, SD = 0.73), the peer comparison intervention (Condition 2) significantly decreased job satisfaction (M = 4.95, SD = 1.48; β = -0.55, 95% CI = [-1.01, -0.09], *P* = 0.021, d = 0.42) and increased burnout (M = 2.47, SD = 0.96; β = 0.33, 95% CI = [0.03, 0.63], *P* = 0.031, d = 0.64). In contrast, PCPs who received leadership support training combined with peer comparison (Condition 3) experienced significantly higher job satisfaction (M = 5.29, SD = 1.27; β = 0.45, 95% CI = [0.02, 0.88], *P* = 0.044, d = 0.25) and lower burnout (M = 2.09, SD = 0.84; β = -0.44, 95% CI = [-0.79, -0.09], *P* = 0.016, d = 0.42) than PCPs who received the peer comparison intervention alone (Condition 2). The results remained statistically significant at the 5% level after a 2-fold Holm-Bonferroni correction that adjusted for multiple hypothesis testing due to simultaneously comparing Conditions 2 v. 1 and Conditions 3 v. 2 (Conditions 2 vs. 1: job satisfaction adjusted *P* = 0.042, burnout adjusted *P* = 0.032; Conditions 3 vs. 2: job satisfaction adjusted *P* = .044, burnout adjusted *P* = 0.032). Finally, we found no significant differences in job satisfaction or burnout between Condition 3 and the Control Condition (*P* = 0.509 and *P* = 0.364, respectively; see Section 10 of the *SI Appendix*).

Robustness Checks & Secondary Analyses. Our aforementioned results about physician well-being were robust to excluding controls for physician characteristics or including the number of positive COVID-19 cases each PCP encountered as a control. Additionally, in a post-hoc placebo test, we confirmed no statistically significant effect of the peer comparison intervention alone or the leadership training on an outcome that we would not expect to be impacted by the interventions (perceived proficiency with the EHR system). Finally, we found no evidence that the negative effects of the peer comparison intervention on well-being were moderated by PCPs' baseline performance. See Section 11 of the *SI Appendix* for these robustness checks and secondary analyses.

Treatment Effect Persistence. To explore the persistence of our interventions' treatment effects, we analyzed survey responses collected four months after the interventions had been discontinued (July 2020; See Figures 2-1 and 2-2 for the regression-estimated treatment effects, and Section 12 of the *SI Appendix* for more details). The negative effect of the peer comparison intervention (Condition 2 vs. Control) on job satisfaction remained significant (Control Condition: M = 5.22, SD = 1.07; Condition 2: M = 4.64, SD = 1.51; β = -0.60, 95% CI = [-1.09, -0.12], *P* = 0.017, d = 0.45). Moreover, PCPs who received leadership support training combined with peer comparison (Condition 3) persistently experienced significantly higher job satisfaction (Condition 3: M = 5.21, SD = 1.38; β = 0.62, 95% CI = [0.14, 1.09], *P* = 0.013, d = 0.39) than PCPs who received the peer comparison intervention alone (Condition 2). These long-term effects on job satisfaction remained significant at the 5% level after a 2-fold Holm-Bonferroni correction (Condition 2 vs. 1: adjusted *P* = 0.026; Conditions 3 vs. 2: adjusted *P* = 0.026). The long-term differences across conditions in burnout were not statistically significant, but they remained directionally consistent with the short-term treatment effects.

Together, these results indicate that the peer comparison intervention negatively impacted two dimensions of physician well-being: job satisfaction and burnout. The harmful effect on job satisfaction lasted for at least four months after the intervention had been discontinued. However, administering the peer comparison intervention with leadership support training appeared to offset these harmful effects.

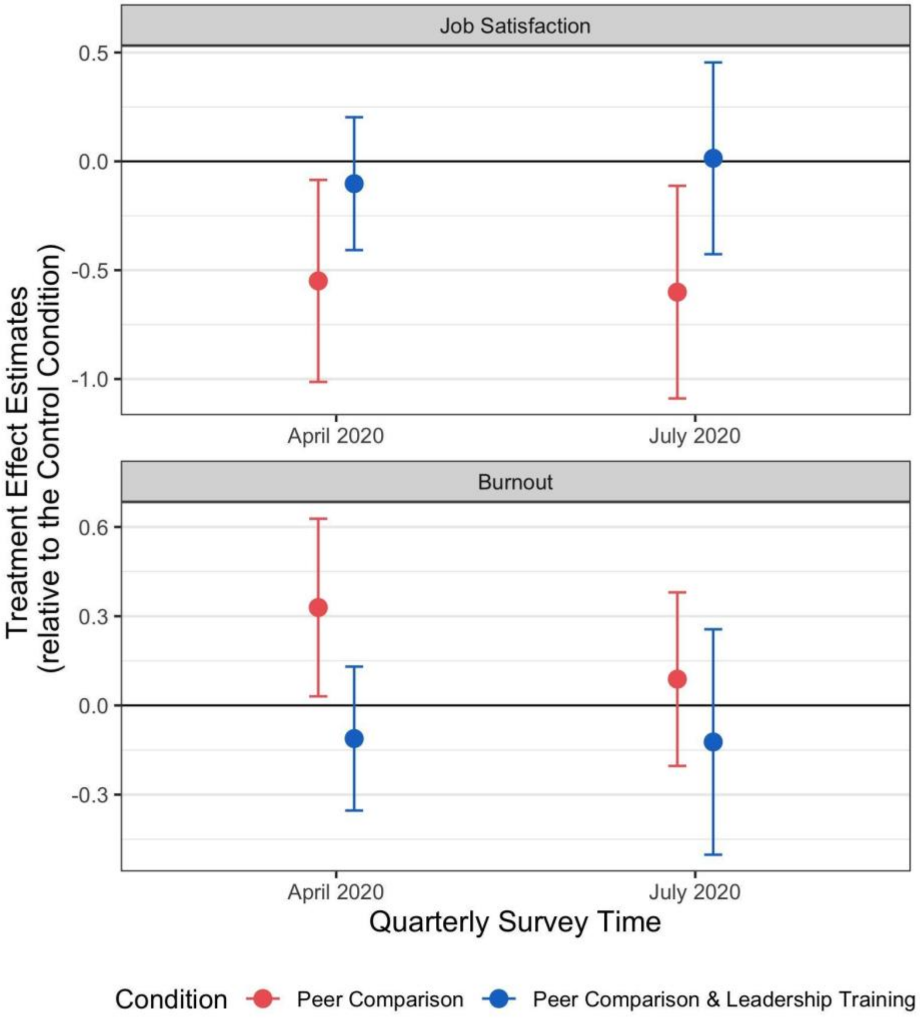


Figure 2-1. Treatment Effect Estimates on Job satisfaction and Burnout.
The blue and red dots reflect the estimated treatment effects of the respective conditions (vs. Control Condition) on job satisfaction (upper panel) and burnout (lower panel). Error bars reflect 95% confidence intervals.

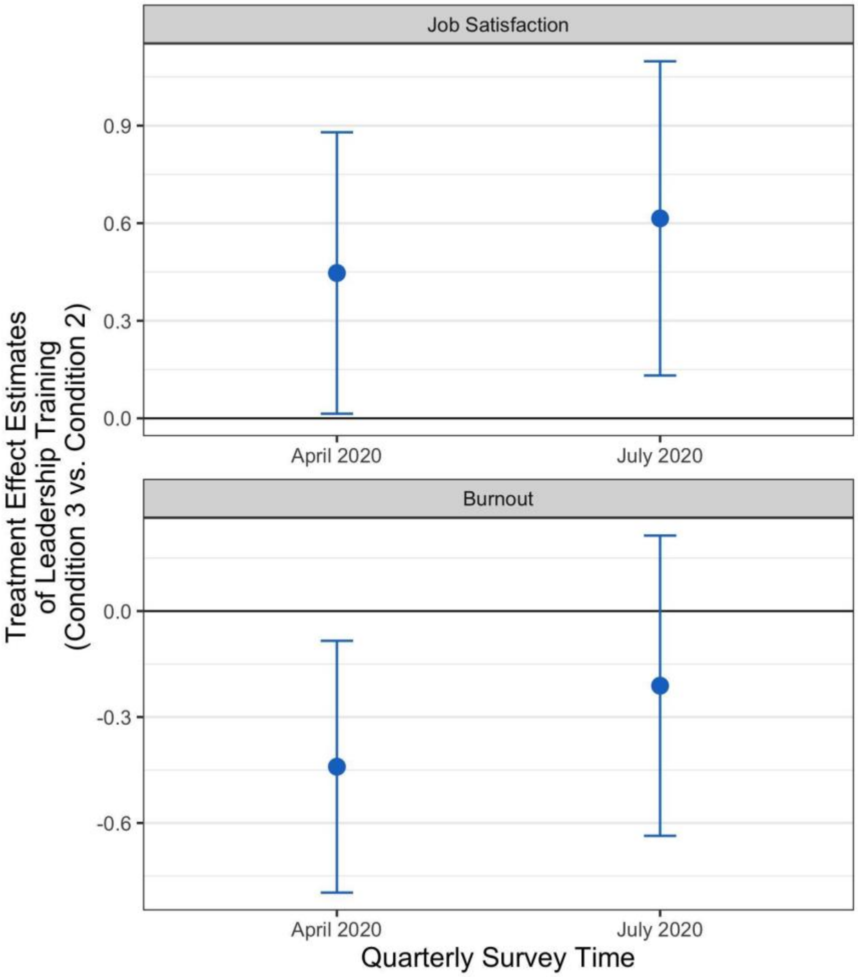


Figure 2-2. Treatment Effect Estimates of Adding Leadership Support Training to the Peer Comparison Intervention.
The blue dots reflect the estimated treatment effects on job satisfaction (upper panel) and burnout (lower panel) of the Peer Comparison and Leadership Training Condition (Condition 3) relative to the Peer Comparison Condition (Condition 2). Error bars reflect 95% confidence intervals.

Exploratory Analysis of Mechanisms

Our finding that training leaders to be more supportive offset the negative effects of the peer comparison intervention on physician well-being led us to investigate one potentially important mechanism. We hypothesized that PCPs may have perceived the administration of the peer comparison intervention alone as signaling a lack of support from leadership (for instance, it may have seemed callous and misdirected). But adding leadership support training may have counteracted this impression. To test this hypothesis, we leveraged a measure of "perceived leadership support" that was included in our quarterly surveys ["I feel supported, understood, and valued by my department leaders" (Richer and Vallerand 1998); 1- "strongly disagree" to 5- "strongly agree"][2].

Figures 2-3 and 2-4 depict the regression-estimated treatment effects of our interventions on perceived leadership support, based on the same regression specification that we used to predict job satisfaction and burnout (see Section 13 of the *SI Appendix*). Compared to PCPs in the Control Condition (April 2020: M = 3.52, SD = 0.91; July 2020: M = 3.46, SD = 0.88), PCPs in the Peer Comparison Condition (Condition 2) reported feeling significantly less supported by their department leaders in both April 2020 (M = 3.02, SD = 1.21; β = -0.60, 95% CI = [-1.06, -0.13], *P* = 0.013, *d* = 0.47) and July 2020 (M = 2.87, SD = 1.21; β = -0.69, 95% CI = [-1.12, -0.26], *P* = 0.002, *d* = 0.56). However, PCPs who received leadership support training combined with peer comparison perceived significantly higher leadership support in April 2020 (Condition 3: M = 3.55, SD = 1.06; β = 0.56, 95% CI = [0.09, 1.03], *P* = 0.021, *d* = 0.47) than PCPs who received the peer comparison intervention alone (Condition 2). This difference is marginally significant in July 2020 (Condition 3: M = 3.38, SD = 1.08; β = 0.49, 95% CI = [0.00, 0.98], *P* = 0.054, *d* = 0.45). Perceived leadership support did not significantly differ between Condition 3 and the Control Condition in April 2020 (*P* = 0.808) or July 2020 (*P* = 0.254). Together, these results are consistent with the interpretation that the peer comparison intervention administered

---

[2] Who would be considered as "department leaders" was deliberately left open to the respondents' interpretation. For example, physician leads may have interpreted "department leaders" as referring to the health system's management, while non-lead physicians may have interpreted it as referring to their physician leads or non-clinical managers.

on its own caused PCPs to feel significantly less supported by their department leaders; but, importantly, leadership support training buffered against this effect.

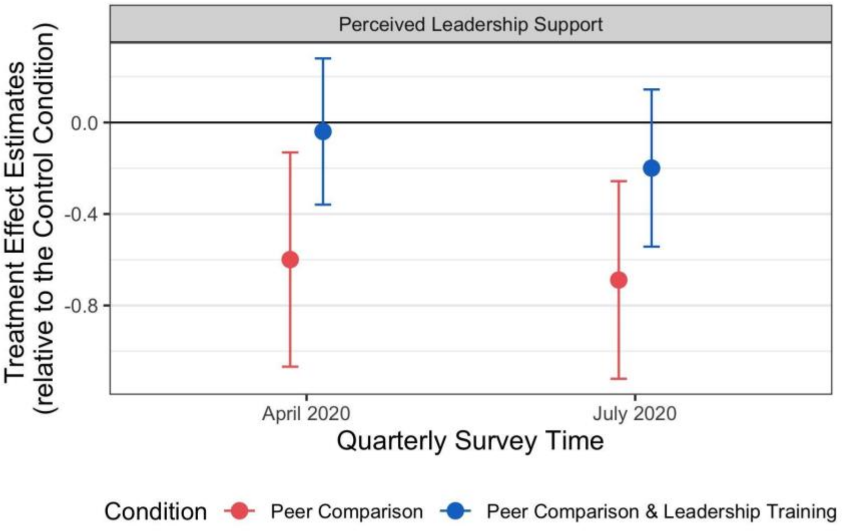


Figure 2-3. Treatment Effect Estimates on Perceived Leadership Support.
The blue and red dots show the estimated treatment effects in the respective conditions (relative to the Control Condition) on perceived leadership support. The error bars reflect 95% confidence intervals.

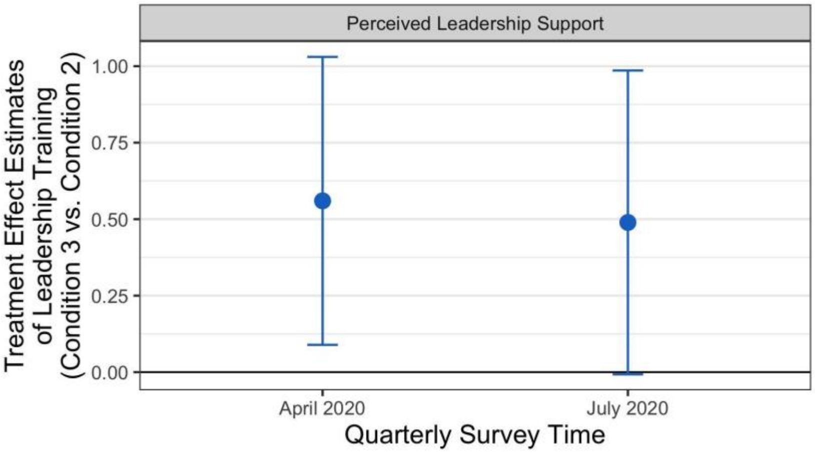


Figure 2-4. Treatment Effect Estimates of Leadership Training on Perceived Leadership Support.
The blue dots show the estimated treatment effects of the Peer Comparison and Leadership Training Condition (Condition 3) relative to the Peer Comparison Condition (Condition 2). Error bars reflect 95% confidence intervals.

To gain further insights into why the peer comparison intervention reduced perceived leadership support, we surveyed PCPs from our study population approximately one year after the intervention had ended (April 2021). Of the original 199 PCPs in the experiment, 169 individuals (85%) were still working for UCLA and were thus invited to take the survey. Of these PCPs, 90.5% (153/169) completed part or all of the survey. Response rates did not significantly differ by the condition PCPs had been assigned to during our experiment ($P = 0.55$ for the F-test of joint significance).

In the survey, we first presented all PCPs (regardless of their experimental condition) with an example of the peer comparison email that had been used in our experiment, and we asked, "Would you prefer that the Department resumes sending these types of emails to physicians?" Of the 150 PCPs who responded to this question, 54% (81/150) preferred that the peer comparison emails not be resumed. More specifically, the proportion of PCPs preferring that peer comparison emails not be resumed was highest (i.e., 68%) among physicians from Condition 2 who had experienced the peer comparison intervention alone (compared to 45% of PCPs from Condition 1 and 50% from Condition 3).

We next asked all PCPs an open-ended question about how receiving such peer comparison emails would make them feel. The open-ended responses again revealed PCPs' negative attitudes towards the peer comparison intervention (see Section 14 of the *SI Appendix* for more details). In particular, these responses suggested two related reasons why the peer comparison intervention would make PCPs feel less supported by leadership, and ultimately, less satisfied with their job and more burned out. First, the leadership's use of peer comparison information in this context was viewed by many PCPs as transgressive. For instance, one PCP stated, “frankly I think it is inappropriate”; another commented that “publicizing data among all faculty feels inappropriate, as if we are all being ranked/valued according to this metric.” Second, leadership's use of one performance metric (the HM completion rate) in the peer comparison emails was viewed by many PCPs as too reductionist. For instance, one PCP stated that the HM completion rates “do not accurately gauge the quality of care a physician provide[s]”; another commented, “top physicians [are] defined by so much more than HM completion.”

Several PCPs explicitly stated that they felt that the peer comparison emails should be accompanied by greater leadership and organizational support. For instance, one PCP cited a lack of “support from upper management to help”; another noted that “completion of health maintenance items should be a ‘system’ effort, not at the individual PCP level.” The leadership support training provided participants (physician leads and non-clinical managers) with information on how completing HM measures would benefit patients, which they were encouraged to share with the non-participating PCPs in their clinics. We conjecture that such information may have helped PCPs—regardless of whether they participated in the training—contextualize the peer comparison emails, making them more amenable to accepting the HM completion metric as a marker of performance, or showing them that management realized that this metric was not the only important measure of job performance. As a result, PCPs may have felt that their leaders were evaluating them more fairly and holistically when leadership support training was included as part of the intervention. Consistent with our speculation, leadership training in Condition 3 did appear to improve perceived leadership support both among physician leads who received the leadership training and among the non-lead physicians who did not personally attend the leadership training (*SI Appendix* Table A2-25), though effects in these subgroups are no longer statistically significant due to the smaller sample sizes.

In sum, these qualitative responses suggest that the manner in which the peer comparison intervention was administered in our context was seen as normatively inappropriate and reductionist; and that adding leadership support training buffered against these perceptions by helping leaders contextualize the intervention.

GENERAL DISCUSSION

Using a five-month field experiment involving 199 physicians and 46,631 patients, we examined the effects of a peer comparison intervention, administered alone or in conjunction with leadership support training, on physicians' preventive care performance and work-related well-being. In this setting, the peer comparison intervention did not significantly improve physicians' performance (measured as order rates for preventive measures). But it did unexpectedly harm job satisfaction and increase burnout, with the effect on job satisfaction persisting for at least four months. Importantly, this negative effect of the peer comparison intervention on physician well-being was substantially attenuated by leadership support training. We find evidence that perceived leadership support may help explain both effects: The peer comparison intervention caused doctors to feel less supported by their leaders, but leadership training buffered against that negative effect.

Although we did not find a statistically significant effect of the peer comparison intervention on physician behavior, previous studies have found significant positive effects, even within similar contexts (Meeker et al. 2016; Navathe et al. 2020; Song et al. 2018). Likewise, peer comparison interventions outside of the healthcare context have had inconsistent effects on targeted behaviors, with some showing null or negative effects (Ashraf, Bandiera, and Lee 2014;

Bandiera, Barankay, and Rasul 2013; Barankay 2012; Buntinx et al. 1993; Bursztyn and Jensen 2015; Hennig-Schmidt, Sadrieh, and Rockenbach 2010) and others showing positive effects (Azmat and Iriberri 2010; Meeker et al. 2016; Navathe et al. 2020; Tran and Zeckhauser 2012; Verbeke, Bagozzi, and Belschak 2016; Vidal and Nossol 2011). There are many different ways to operationalize and communicate peer comparison interventions. We speculate that details in the intervention design, implementation, and context matter in determining their success. Among other aspects of our design, publicly displaying a list of the Top 25 performers using a composite performance metric may have curbed any motivating effects of peer comparisons for a few reasons. First, PCPs may have found it reductionist for their leaders to evaluate their job performance using a single metric (Ranganathan and Benson 2020). Second, it may have seemed unjust to evaluate performance in relative terms (i.e., Top 25), rather than using an absolute criterion that reflects top quality of care (Mayer, Venkatesh, and Berwick 2021). Using an absolute criterion instead would have also allowed for the public list of top performers to potentially grow over time, which could have motivated people by highlighting a growing trend (Sparkman and Walton 2017). Third, highlighting exemplary performance (e.g., Top 25 Physicians) could be discouraging to people who do not believe improvement is possible (Lockwood and Kunda 1997). In our case, people at the bottom of the performance distribution were the most likely to feel incapable of behavior change, even though they had the most room for improvement (see Section 15 of the *SI Appendix* for details). These features of our design may have been perceived as particularly inappropriate or offensive in the present social context, where physicians' roles and responsibilities typically involve communal norms that foster care and collaboration (Blader, Gartenberg, and Prat 2020; Gallus et al. 2021).

Our findings offer three key contributions to the peer comparison literature. First, we provide the first field experimental evidence to our knowledge of the negative effects of a peer comparison intervention on workers' job satisfaction and burnout. Second, our findings underscore the importance of attending to the way in which implementation details of a peer comparison intervention are perceived by targeted individuals within the relevant social context. Researchers have recently argued that behavioral interventions are not experienced "in a vacuum," but rather that they are "embedded in a social ecosystem involving an implicit or explicit interaction between targeted individuals and the [intervention] designer" (Fox et al. 2020). According to this account, people attend to the details of behavioral interventions—especially interventions that have been newly introduced—to infer their leaders' beliefs and values. When such inferences are negative (e.g., *my leaders do not seem to support me*), targeted individuals may respond unfavorably to the intervention. Thus, to enhance the effectiveness of behavioral interventions, our research suggests that policymakers and organizational leaders ought to engage targeted individuals in the *design phase* of an intervention, probe the inferences they draw about it, and revise the design to reduce negative inferences before scaling the intervention in the field. Finally, our work highlights that when leaders offer the necessary context and support to accompany a peer comparison intervention, recipients may draw more positive inferences about their leaders' intent. This can buffer against the harmful effects of peer comparison interventions on well-being.

Our study has several limitations that suggest interesting directions for future research. First, our interventions had to be discontinued after only five months due to the COVID-19 pandemic. It remains an open question whether the peer comparison intervention would have become normalized over time and thus might have stopped affecting physician well-being.

Second, the leadership support training intervention was multifaceted with a variety of components and a broad curriculum. Future research is needed to discern which aspects of the leadership support training affected job satisfaction and burnout. Finally, although job satisfaction and burnout were pre-registered secondary outcomes, we did not predict a negative effect *a priori*. It would be valuable to design future experiments to deductively test hypotheses concerning the conditions under which a broader range of behavioral interventions harm the well-being of targeted individuals.

When measuring both the behavioral and psychological impact of an intervention, difficult trade-offs may arise: How are we to decide whether an intervention is worthwhile if it produces desired behavior change (e.g., motivating physicians to improve patient outcomes) but reduces well-being? For instance, notifying doctors about their patients who suffered fatal overdoses has been shown to reduce subsequent opioid prescriptions (Doctor et al. 2018). Although such notifications were likely highly aversive to doctors, one could argue that this is justified by the behavior change that saves lives. Naturally, other cases will be more ambiguous. In order to design and deploy interventions that holistically improve social welfare, researchers, policymakers, and ethicists will need to continue examining these trade-offs and develop new approaches to quantify or even price the psychological consequences of interventions (Allcott and Kessler 2019; Butera et al. 2022).

CONCLUSION

Behavioral interventions such as providing peer comparison information offer attractive, cost-effective ways to promote positive behavior change. Our work suggests that if policymakers and organizational leaders only measure the behavioral outcomes of such interventions, they risk overlooking important effects on less visible outcomes, such as job satisfaction and burnout.

These psychological outcomes need to be accounted for to estimate the aggregate impacts of policies and to improve their design and implementation.

MATERIALS AND METHODS

Setting. Between November 5, 2019 and March 3, 2020, we collaborated with the UCLA Health Department of Medicine (DOM) Quality Team to run a field experiment across the health system's entire primary care network. In line with the DOM Quality Team's goal of motivating physicians to improve their patients' uptake of preventive care services, all PCPs in our study were part of a pay-for-performance program that incentivized them to meet a threshold HM completion rate. For each PCP, the HM completion rate reflects the proportion of recommended preventive care measures that were completed by their patients in a given time period. There are 26 different measures recommended by the U.S. Preventive Service Task Force and other medical associations (e.g. American Diabetes Association), of which the DOM Quality Team identified nine high-priority "focus measures" (e.g., diabetes hemoglobin A1c screening). Details regarding how HM completion rates were calculated are available in Section 1 of the *SI Appendix*.

Experimental Design. The experiment was originally designed and pre-registered to span twelve months but was discontinued in March 2020 due to the COVID-19 pandemic (ClinicalTrials.gov # NCT04237883). The experiment included 199 PCPs across 42 clinic sites that specialized in internal medicine, geriatrics, or family medicine, and that had a clinical full-time employment (FTE) rate of at least 50%. PCPs were unaware of this research investigation. They were cluster-randomized at the clinic level to one of three study conditions: Control, Peer Comparison, Peer Comparison and Leadership Training (Table 2-1). Each condition involved 14 clinics. For more information on the inclusion criteria and randomization algorithm, see Sections

1 and 2 in the *SI Appendix*. Section 3 in the *SI Appendix* shows that conditions were balanced on all observable patient, physician, and clinic characteristics.

All PCPs received monthly emails from the DOM Quality Team that informed them of their HM completion rate over the prior three months. They were signed by the health system's management. The emails contained other information and links intended to help PCPs improve HM completion rates (see *SI Appendix* Sections 4 and 5 for email details, examples, and email engagement statistics). Emails were sent near the start of each month. A maximum of two reminder emails–identical to the initial email–were sent to those who had not opened the initial email after 7 and 14 days, respectively.

Table 2-1. Descriptions of Intervention(s) Implemented in Each Condition

Condition	Main Intervention Elements
1. Control	- Monthly emails informed PCPs of their HM completion rate over the prior three months, the focus measure on which they had performed the best, and the two focus measures that they could most improve on
2. Peer Comparison	- Same information as in the monthly emails in the Control Condition - Monthly emails also included list of the names of the top 25 PCPs as well as messaging based on the recipient's placement in the performance distribution (Top 25 Physician, High Performer, Almost High Performer, Low Performer)
3. Peer Comparison and Leadership Training	- Same monthly emails as in the Peer Comparison Condition - Clinical physician leaders and non-clinical managers received two training workshops (on how to provide effective support to fellow physicians) and monthly check-in emails - Physician leads had one-on-one meetings with members of the DOM Quality Team to identify specific challenges at their clinics and brainstorm strategies to address these challenges

For PCPs in the Peer Comparison Condition (Condition 2) and the Peer Comparison and Leadership Training Condition (Condition 3), the emails also included information about their peers' performance. These emails contained a banner displaying the names of PCPs whose HM completion rate in the prior three months was within the top 25 of all PCPs in the study population. These PCPs were labeled "Top 25 Primary Care Physicians." Additionally, emails in Conditions 2 and 3 informed PCPs of their relative standing in terms of HM completion rates compared to all other PCPs in the prior three months.

- PCPs who were one of the Top 25 PCPs in a given month received a message saying, "Congratulations! You are a Top 25 Primary Care Physician in [respective month]!"
- PCPs whose HM completion rate was above 65% but who were not one of that month's Top 25 PCPs were informed, "Congratulations! You are a High Performer!"
- PCPs whose HM completion rate was between 55%-65% were told, "You are almost a High Performer."
- PCPs with an HM completion rate under 55% were informed that "the majority of physicians have a HM completion rate of 55% or higher."

The emails further informed all PCPs of the HM completion rates necessary to be a "High Performer" or a "Top 25 Primary Care Physician," whichever was more proximate, and they encouraged PCPs to improve their performance (or maintain their performance if they were already a Top 25 PCP). The performance tier cutoffs had been selected to ensure that most PCPs would fall into the two middle performance tiers, where they would feel close to reaching the next-higher group (see Section 4 of the *SI Appendix* for details).

For the 14 clinics assigned to the Peer Comparison and Leadership Training Condition (Condition 3), physician leads and non-clinical managers participated in two four-hour training workshops, one in December 2019 and one in March 2020. These workshops focused on training attendees to develop their leadership skills and effectively support their fellow PCPs.

Importantly, a primary goal of the training was to help attendees provide fellow PCPs at their clinics with the necessary contextual information to understand and appreciate why UCLA Health uses HM completion rates to measure performance. Among the physician leads in Condition 3 clinics, 11 were in our experiment, overseeing a total of 59 other PCPs.

Following the training workshops, clinic physician leads and their non-clinical manager counterparts received additional resources from the DOM Quality Team through monthly check-in emails. Additionally, in-person meetings with clinic physician leads in Condition 3 occurred in January and February 2020. During these one-on-one meetings, the DOM Quality Team helped physician leads identify specific challenges at their clinic and develop corresponding solutions. Section 6 of the *SI Appendix* includes detailed information about the leadership training intervention along with materials from the training workshops.

Data. For each PCP, we measure their order rate for patients who satisfy the following pre-registered criteria: 1) patients were empaneled to a PCP participating in the field experiment (based on the attribution logic laid out in Section 1 of the *SI Appendix*), 2) had at least one in-office visit with their PCP during the intervention period (November 5, 2019 to March 3, 2020), and 3) had at least one focus measure due at the time of that in-office visit. A total of 46,631 patients met these inclusion criteria. See Section 3 of the *SI Appendix* for more information on the sample characteristics.

From October 2019 through July 2020, PCPs who were part of the field experiment were asked to complete quarterly surveys assessing their experiences at work and participation in professional activities. The surveys (sent by the DOM leadership) collected longitudinal measures of job satisfaction, burnout, and feelings of leadership support, along with other measures not pertinent to the current investigation. Since the field experiment had to be discontinued in March 2020 (due to COVID-19), we used the April 2020 survey data for our primary analysis. We also examined the sustained impact of our interventions by analyzing the July 2020 survey data. Completion of these surveys was tied to the aforementioned pay-for-performance incentive program. Thus, 93.0% (185/199) of physicians completed the April 2020 survey and 88.4% (176/199) completed the July 2020 survey (see *SI Appendix* Section 7 for survey details and quarterly completion rates). See Figure 2-5 for a timeline of the study.

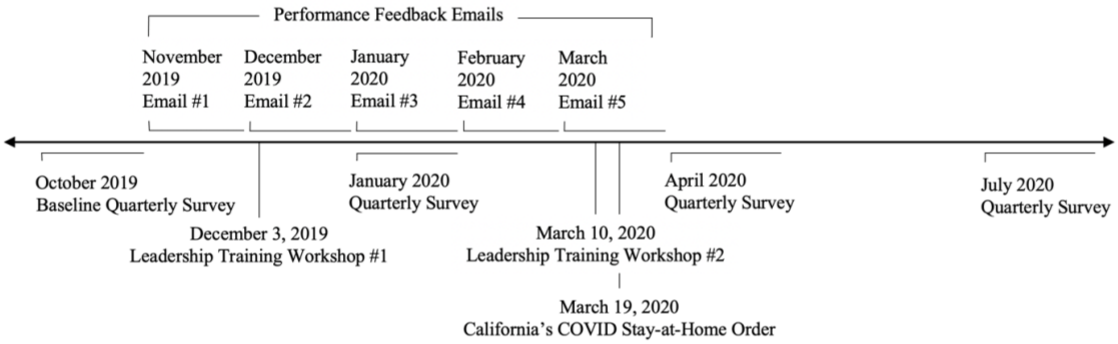


Figure 2-5. Study Timeline.
This timeline depicts the timing of the relevant events in the study. The L-shaped lines depict events that occurred over sustained periods of time: the performance feedback emails were initially sent at the beginning of each month and up to two reminders were sent during the month to those who had not opened the initial emails; PCPs had approximately two weeks to complete the quarterly surveys. The straight vertical lines depict discrete events. For ease of visualization, the email and survey dates are approximate. See Section 4 of the *SI Appendix* for the precise dates of each email sent and survey launched.

Measures. For each patient empaneled to a PCP, our pre-registered primary behavioral outcome was the HM order rate for focus measures that were due at the patient's first in-office primary care visit during the intervention period (hereafter, "order rate"). It equals the share of open HM focus measures (i.e., focus measures recommended for the patient based on the national guidelines but not yet completed at the time of the patient's first visit) that were ordered by the PCP within seven days following the patient's first visit:

$$\text{Order Rate} = \frac{Number\ of\ ordered\ HM\ focus\ measures\ 7\ days\ following\ first\ visit\ during\ the\ study\ period}{Number\ of\ open\ HM\ focus\ measures\ at\ the\ time\ of\ the\ first\ visit\ during\ the\ study\ period}$$

The order rate was chosen as the primary behavioral outcome because it is clinically important and not subject to factors outside the PCPs' control (e.g., patients' willingness or ability to obtain preventive service).

Our pre-registered secondary outcomes included two measures of physician well-being: job satisfaction and burnout, which we assessed using validated single-item scales in every quarterly survey. Job satisfaction was measured with the question, "Taking everything into consideration, how do you feel about your job as a whole?", with responses ranging from "extremely dissatisfied" to "extremely satisfied" on a seven-point Likert scale (Dolbier et al. 2005). We used a validated and widely-used burnout measure (Dolan et al. 2015): "Overall, based on your definition of burnout, how would you rate your level of burnout?", with five response options ranging from 1 = "I enjoy my work. I have no symptoms of burnout" to 5 = "I feel completely burned out and often wonder if I can go on. I am at the point where I may need some changes or may need to seek some sort of help."

Statistical Analysis. To compare patient-level order rates between conditions, we estimated a mixed effects binomial logistic regression model. This model assumes that each patient's number of orders placed follows a binomial distribution, where the number of trials is the patient's number of open topics and a logit-linear function is used to estimate the probability that a patient has an order placed for any given open topic. Physician and clinic random effects account for clustering of patients. The pre-registered baseline controls are: patient characteristics, including their completion rate measured from July-October 2019, age, gender, and zip code (using fixed effects for the 3-digit zip code for all Southern California zip codes and a single indicator for everyone else)[3]; and physician characteristics, including their gender, race, years since graduating medical school, and years of working at UCLA Health. We pre-registered the following gatekeeping approach for our analysis in order to reduce multiple hypothesis testing (Dmitrienko and Tamhane 2007): We would first test whether HM order rates differed between the combination of Conditions 2 and 3 versus Condition 1. If and only if this comparison was statistically significant, we would conduct additional comparisons across conditions using a Holm-Bonferroni adjustment, for an overall significance level of 0.05. Our results are robust to alternate specifications including binomial logistic regressions with standard errors clustered at the clinic level, mixed effects linear models with physician and clinic random effects, and linear regression models with standard errors clustered at the clinic level (reported in Section 8 of the *SI Appendix*).

To assess differences in survey measures (e.g., job satisfaction, burnout) between conditions, we used linear regression models, with cluster-robust standard errors at the clinic level. These regressions controlled for the respective outcome measure taken from the baseline October 2019 quarterly survey, as well as the same set of physician demographics as pre-registered in our analysis of order rates (physician gender, race, years since graduating medical school, and years of working at UCLA Health).

---

[3] We had also pre-registered controlling for patient comorbidity and insurance plan. We unexpectedly did not have access to these variables, and thus, they are not included in the reported regressions.

CHAPTER 2

APPENDIX

1. HM Completion Rate and UCLA Patient Attribution Model Details

HM Completion Rate Details

PCPs' Health Maintenance (HM) completion rate equals the total completed primary care clinical quality measures divided by the total number of open clinical quality measure opportunities within their patient panel. There are a total of 26 clinical quality measures (focus and complementary measures) tracked by the UCLA Department of Medicine's (DOM) Quality Team. The measures are based on the recommendations of the United States Preventive Services Taskforce. The HM completion rate is used, in conjunction with other measures (e.g., productivity, patient satisfaction), to determine each PCP's incentive compensation.

Focus measures:
Breast Cancer Screening: Mammogram
Cervical Cancer Screening
Colorectal Cancer Screening
Diabetic Eye Exams
Diabetic Foot Exams
Diabetes HbA1c Testing
Diabetes Nephropathy Testing
Chlamydia Screening (Med-Peds/FM only*)
HPV Immunization (Med-Peds/FM only*)

Complementary measures:
Abdominal Aortic Aneurysm (AAA) Screening
Annual Preventive Wellness Visit
Diabetes: Pneumococcal Vaccine
DTaP/Tdap/Td Vaccine
Complete Hepatitis A Vaccines
Hepatitis B Vaccines
Hepatitis C Screening
HIV Screening
IPV Vaccines
Meningococcal Vaccine (MCV4)
MMR Vaccines
Osteoporosis Early Detection DEXA Scan
Pneumococcal Vaccine
Shingles (Shingrix) Vaccine
Statin prescribed for ASCVD Prevention or Treatment

Tdap During Pregnancy (If > 28 Weeks)
Tdap/Td Vaccine
*Note: Med-Peds refers to Internal Medicine-Pediatrics. FM refers to Family Medicine.

UCLA Health Primary Care Patient Attribution Model Details

The UCLA Health Primary Care Patient Attribution Model is the methodology used to designate patients to each PCP's patient panel. The attribution model is detailed below:

- If a patient has seen the PCP listed in UCLA's electronic health record (Epic Systems ©1979) in the prior 3 years, the patient is attributed to that provider.
- If the patient has not seen the PCP listed in the electronic health record in the prior 3 years or if the electronic health record's PCP field is blank or if the provider listed in the CareConnect PCP field is a UCLA specialist, then the patient's visit history over the prior 3 years is reviewed and the UCLA PCP is attributed as follows:
    1) The UCLA PCP with a preventive/wellness visit in the prior 1 year is attributed first
    2) If there is no preventive/wellness visit in the prior 1 year, the UCLA PCP with the highest volume of visits is attributed
    3) If there is a tie in either the preventive/wellness visit or volume of visit scenario, the UCLA PCP with the most recent visit is attributed.

2. Inclusion Criteria, Randomization Algorithm, and Pre-registration Details

Physician Inclusion Criteria

PCPs were included in the experiment if they satisfied the following criteria at the beginning of the intervention period (i.e., in October 2019):

1) They were part of the UCLA Health DOM primary care network
2) They were a Board-certified Internal Medicine, Geriatrics, Internal Medicine-Pediatrics, and/or Family Medicine physician
3) They had a clinical full-time employment (FTE) level of at least 50% (for reference, 100% FTE is equivalent to 40 hours of clinical work per week)
4) They were eligible for a quarterly primary care quality incentive based on meeting DOM's productivity threshold, and
5) They had a panel size of over 50 patients.

PCPs included in our experiment accounted for 83% of all regularly working PCPs (i.e., with at least 50% FTE) in the UCLA Health DOM network.

Randomization Algorithm

Randomization was performed at the clinic level, using a 1:1:1 allocation ratio. Clinics were stratified by UCLA clinic group (DOM vs. CPN/EIMG). For reference, DOM refers to the Department of Medicine clinic group; and CPN/EIMG stands for Community Physician Network/Entertainment Industry Medical Group, and it is treated as a single clinic group by UCLA Health. A covariate-constrained randomization procedure was used to randomize clinics within UCLA clinic groups. This involved 1) generating 100,000 random allocations, 2) computing a balance score for each allocation, and 3) randomly drawing one from the 1,000 most balanced allocations as our implemented allocation. Factors incorporated into the balance score were 1) total clinical FTE, and 2) clinic-level baseline HM completion rates. Since clinics were being randomized between 3 arms, we used one-way ANOVA F-statistics (evaluating differences in each factor across arms) to measure imbalance, and then computed a balance score by summing the F-statistics for the two factors. Randomization was performed using r.

Pre-Registration Details

The pre-registration document can be found on Clinicaltrials.gov (ClinicalTrials.gov Identifier: NCT04237883).

Our pre-registration was submitted after the experiment started because we had to launch at a specific date based on external deadlines set by UCLA Health before we were able to put together a detailed analysis plan. Importantly, we did not have access to data from the experimental period prior to our pre-registration submission.

3. Sample Characteristics

Table A2-1. Sample Characteristics

	Condition 1	Condition 2	Condition 3
	Control	Peer Comparison	Peer Comparison and Leadership Training
Clinic Characteristics	(N=14)	(N=14)	(N=14)
Clinical Full Time Employment (FTE)	4.31	4.33	4.43
Baseline HM Completion Rate	0.52	0.52	0.54
CPN/EIMG, n (%)	8 (57%)	7 (50%)	7 (50%)
Physician Characteristics	(N=65)	(N=64)	(N=70)
Gender, n (%)			
Male	27 (42%)	22 (34%)	27 (39%)
Female	33 (51%)	35 (55%)	30 (43%)
Unknown	5 (8%)	7 (11%)	13 (19%)
Race, n (%)			
White	30 (46%)	29 (45%)	33 (47%)
Black	1 (2%)	2 (3%)	0
Asian	20 (31%)	16 (25%)	16 (23%)
Native Hawaiian/Pacific Islander	0	0	2 (3%)
Other	6 (9%)	4 (6%)	3 (4%)
Multiple	1 (2%)	0	2 (3%)
Unknown	7 (11%)	13 (20%)	14 (20%)
Patient Panel Size, Mean (SD)	1507 (808)	1427 (812)	1501 (762)
Years at UCLA, Mean (SD)	7.0 (6.5)	6.2 (6.1)	4.7 (3.6)
Baseline Job Satisfaction* (October 2019), Mean (SD)	5.27 (1.03)	5.40 (1.17)	5.54 (1.26)
Baseline Burnout* (October 2019), Mean (SD)	2.13 (0.72)	2.44 (0.90)	2.37 (1.00)
Patient Characteristics	(N=16,425)	(N=14,781)	(N=15,425)
Age (in years) at Visit, Mean (SD)	53.4 (16.6)	52.5 (17.4)	52.8 (16.8)
Gender, n (%)			
Female	11,938 (72.7%)	10,951 (74.1%)	11,125 (72.1%)
Male	4,487 (27.3%)	3,830 (25.9%)	4,300 (27.9%)
Baseline HM Completion Rate, Mean (SD)	0.28 (0.30)	0.29 (0.30)	0.27 (0.30)

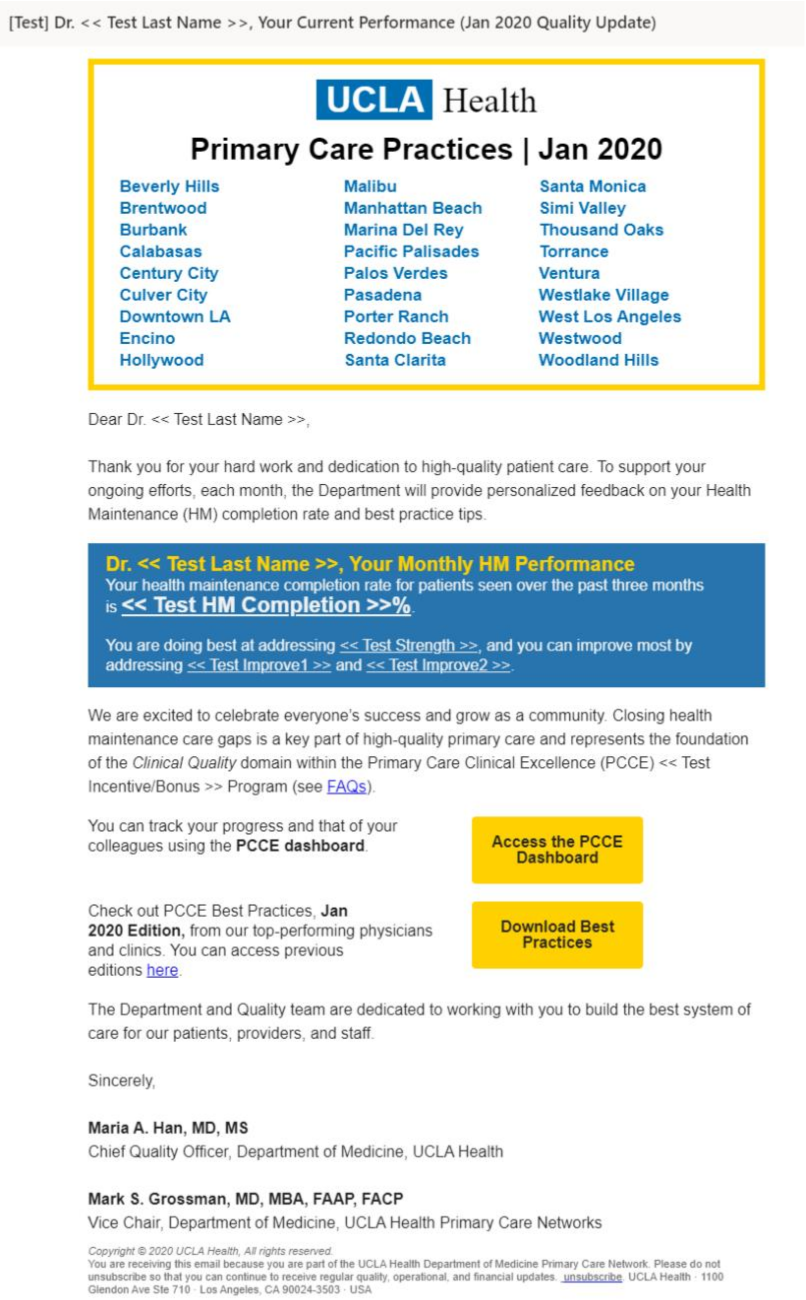
Note: This table displays clinic, physician, and patient-level characteristics across the three study conditions. *An F-test of joint significance confirms that the conditions were balanced during the baseline period in job satisfaction (p = 0.432) and burnout (p = 0.134).

4. Monthly Performance Feedback Email Details

In all conditions, PCPs received monthly emails from the DOM Quality Team informing them of their HM completion rate over the prior three months, the focus measure on which they had performed the best, and the two focus measures on which they had performed the worst. All emails contained links to: 1) a dashboard showing their performance on all nine focus measures, 2) a document that was updated monthly with tips and guidance for improving performance on focus measures, and 3) a document containing frequently asked questions about the DOM's pay-for-performance program. See below for email examples, images of these resources, and engagement statistics including email open rates. Emails were sent near the beginning of each month. A maximum of two reminder emails, which were identical to the initial email, were sent to those who had not opened the initial email after 7 and 14 days, respectively.

Monthly Performance Feedback Email (Condition 1)

This is an example template of monthly performance feedback emails for physicians in Condition 1 of our field experiment. Text within brackets (<<text>>) was personalized for each physician. The email contained the following hyperlinks (see Section 5 for examples): FAQ sheet regarding the PCCE Program, Tableau Dashboard which provided a detailed breakdown of the physician's quality measure performance, and a PDF of the monthly Best Practices document for both the current month and the previous months.



[Test] Dr. << Test Last Name >>, Your Current Performance (Jan 2020 Quality Update)

UCLA Health

Primary Care Practices | Jan 2020

Beverly Hills
Brentwood
Burbank
Calabasas
Century City
Culver City
Downtown LA
Encino
Hollywood
Malibu
Manhattan Beach
Marina Del Rey
Pacific Palisades
Palos Verdes
Pasadena
Porter Ranch
Redondo Beach
Santa Clarita
Santa Monica
Simi Valley
Thousand Oaks
Torrance
Ventura
Westlake Village
West Los Angeles
Westwood
Woodland Hills

Dear Dr. << Test Last Name >>,

Thank you for your hard work and dedication to high-quality patient care. To support your ongoing efforts, each month, the Department will provide personalized feedback on your Health Maintenance (HM) completion rate and best practice tips.

Dr. << Test Last Name >>, Your Monthly HM Performance

Your health maintenance completion rate for patients seen over the past three months is **<< Test HM Completion >>%**.

You are doing best at addressing << Test Strength >>, and you can improve most by addressing << Test Improve1 >> and << Test Improve2 >>.

We are excited to celebrate everyone's success and grow as a community. Closing health maintenance care gaps is a key part of high-quality primary care and represents the foundation of the *Clinical Quality* domain within the Primary Care Clinical Excellence (PCCE) << Test Incentive/Bonus >> Program (see FAQs).

You can track your progress and that of your colleagues using the **PCCE dashboard**.

Check out PCCE Best Practices, **Jan 2020 Edition,** from our top-performing physicians and clinics. You can access previous editions here.

The Department and Quality team are dedicated to working with you to build the best system of care for our patients, providers, and staff.

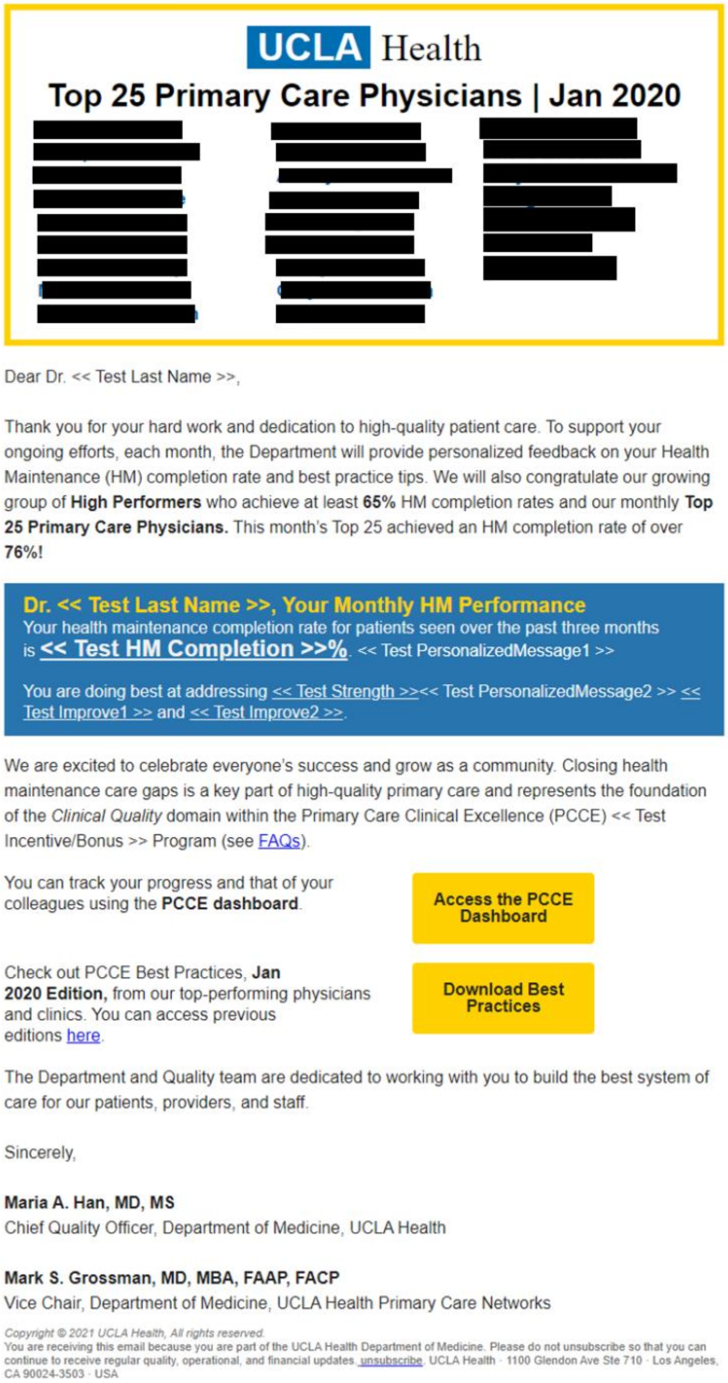
Sincerely,

Maria A. Han, MD, MS
Chief Quality Officer, Department of Medicine, UCLA Health

Mark S. Grossman, MD, MBA, FAAP, FACP
Vice Chair, Department of Medicine, UCLA Health Primary Care Networks

Monthly Performance Feedback Email with Peer Comparison Information (Conditions 2 & 3)

This is an example template of monthly performance feedback and peer comparison emails for physicians in Conditions 2 and 3 of our field experiment. All the hyperlinks in Conditions 2 and 3 emails were the same as those in Control 1 emails. Also, the email layout and design (including style, length, and non-experimental content) were crafted to be as similar as possible between these two conditions and Condition 1.



UCLA Health

Top 25 Primary Care Physicians | Jan 2020

[redacted]

[redacted]

[redacted]

Dear Dr. << Test Last Name >>,

Thank you for your hard work and dedication to high-quality patient care. To support your ongoing efforts, each month, the Department will provide personalized feedback on your Health Maintenance (HM) completion rate and best practice tips. We will also congratulate our growing group of **High Performers** who achieve at least **65%** HM completion rates and our monthly **Top 25 Primary Care Physicians.** This month's Top 25 achieved an HM completion rate of over **76%!**

Dr. << Test Last Name >>, Your Monthly HM Performance
Your health maintenance completion rate for patients seen over the past three months is **<< Test HM Completion >>%**. << Test PersonalizedMessage1 >>

You are doing best at addressing << Test Strength >><< Test PersonalizedMessage2 >> << Test Improve1 >> and << Test Improve2 >>.

We are excited to celebrate everyone's success and grow as a community. Closing health maintenance care gaps is a key part of high-quality primary care and represents the foundation of the *Clinical Quality* domain within the Primary Care Clinical Excellence (PCCE) << Test Incentive/Bonus >> Program (see FAQs).

You can track your progress and that of your colleagues using the **PCCE dashboard**.

Check out PCCE Best Practices, **Jan 2020 Edition,** from our top-performing physicians and clinics. You can access previous editions here.

The Department and Quality team are dedicated to working with you to build the best system of care for our patients, providers, and staff.

Sincerely,

Maria A. Han, MD, MS
Chief Quality Officer, Department of Medicine, UCLA Health

Mark S. Grossman, MD, MBA, FAAP, FACP
Vice Chair, Department of Medicine, UCLA Health Primary Care Networks

The following information was disp"ayed'In the emails in Conditions 2 and 3.

- At the top of the email, the names of the Top 25 PCPs were listed in a banner. This list was updated each month.
- The first paragraph included a high performer benchmark (65% completion rate). This benchmark was held constant during the study period.

- A personalized message notified PCPs about how they compared to other physicians. Physicians were classified into one of four performance tiers, and the personalized message varied depending on their classification. The subject line also varied depending on their classification. Below we provide more information on the four performance tiers:

    - **Top 25 Performer:** Participants were labelled a “Top 25 Primary Care Physician” in a given month if their 90-day HM completion rates were among the top 25 scores across all three study conditions. The email subject line and the email body congratulated them on being a Top 25 PCP.

    - **High Performer:** Participants who achieved a 90-day HM completion rate of 65% or higher, but were not among the Top 25, were labelled as a “High Performer”. A 65% threshold was chosen for the High Performer threshold to be above the median HM completion rate. At baseline, the 65% threshold corresponded with the 59th percentile across all PCPs in our study. The email subject line and the email body congratulated them on being a High Performer and encouraged them to become a Top 25 performer.

    - **Almost High Performer:** Participants with a 90-day HM completion rate between 55% and 65% were labelled “Almost a High Performer”. At baseline, 55% and 65% HM completion rates corresponded with the 29th percentile and the 59th percentile, respectively. The email subject line and the email body both acknowledged their status as almost being a High Performer and encouraged them to become a High Performer.

    - **Low Performer:** Participants with a 90-day HM completion rate lower than 55% were internally classified as “Low Performers”. However, to avoid offending these physicians, this negative label was not mentioned in the emails. Their email subject line was instead worded, “Your Current Performance” and the personalized message in the email body noted that, “The majority of physicians have an HM completion rate of 55% or greater”.

Email Distribution Schedule

The emails were distributed monthly, with two reminder emails per month for those who had not yet opened that month's email. Email operations were conducted using Mailchimp©. The date each email was sent out is listed below:

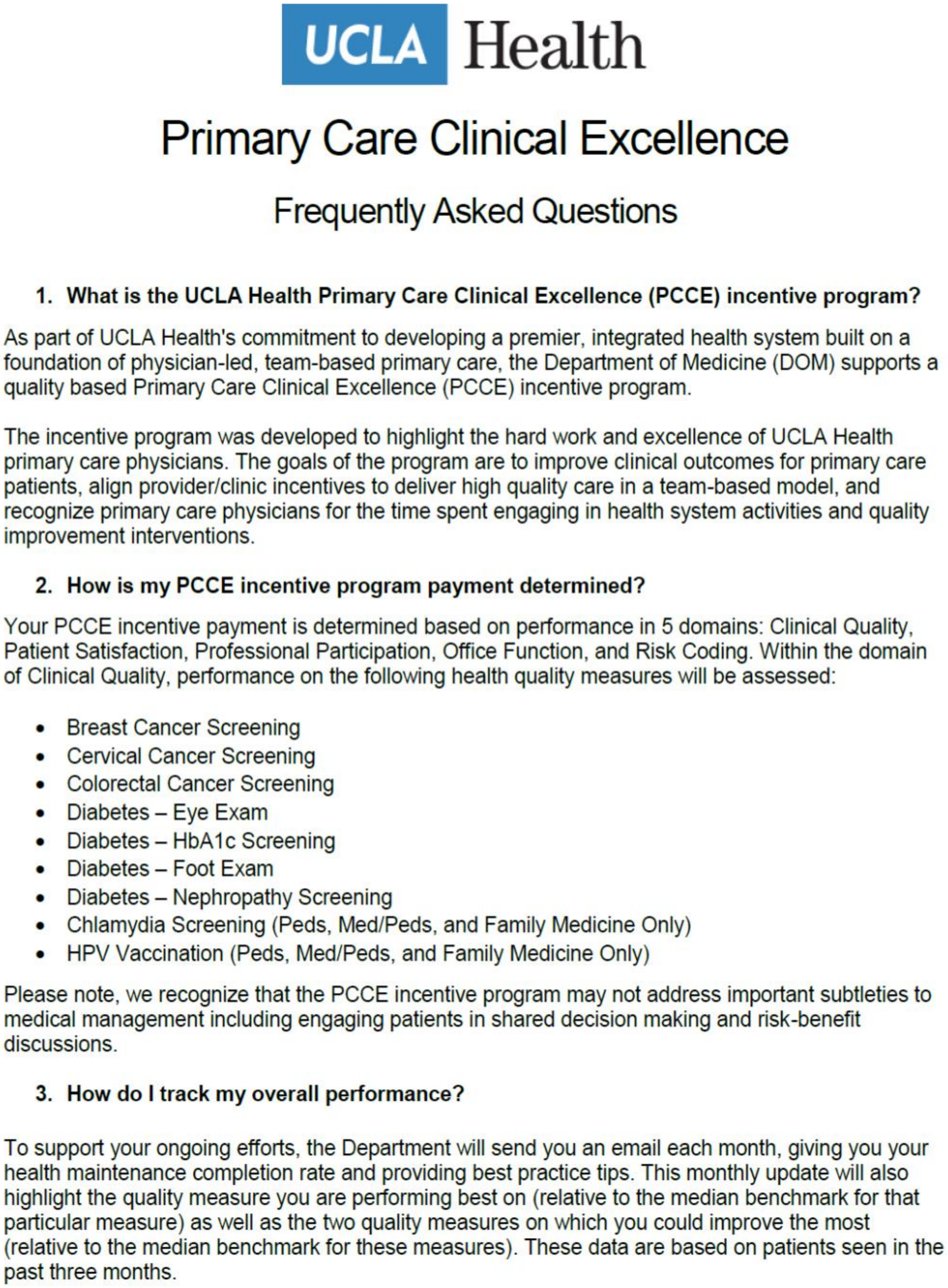
Table A2-2. Email Distribution Dates

Monthly Email	**1st Follow Up Email**	**2nd Follow Up Email**
November 5th, 2019	November 12th, 2019	November 18th, 2019
December 4th, 2019	December 11th, 2019	December 18th, 2019
January 16th, 2020	January 23rd, 2020	January 30th, 2020
February 11th, 2020	February 18th, 2020	February 25th, 2020
March 4th, 2020	March 10th, 2020	March 17th, 2020

5. Monthly Email Materials and Engagement Statistics

FAQ Document

This FAQ document was provided as a reference in each of the monthly emails.



UCLA Health

Primary Care Clinical Excellence

Frequently Asked Questions

1. What is the UCLA Health Primary Care Clinical Excellence (PCCE) incentive program?

As part of UCLA Health's commitment to developing a premier, integrated health system built on a foundation of physician-led, team-based primary care, the Department of Medicine (DOM) supports a quality based Primary Care Clinical Excellence (PCCE) incentive program.

The incentive program was developed to highlight the hard work and excellence of UCLA Health primary care physicians. The goals of the program are to improve clinical outcomes for primary care patients, align provider/clinic incentives to deliver high quality care in a team-based model, and recognize primary care physicians for the time spent engaging in health system activities and quality improvement interventions.

2. How is my PCCE incentive program payment determined?

Your PCCE incentive payment is determined based on performance in 5 domains: Clinical Quality, Patient Satisfaction, Professional Participation, Office Function, and Risk Coding. Within the domain of Clinical Quality, performance on the following health quality measures will be assessed:

- Breast Cancer Screening
- Cervical Cancer Screening
- Colorectal Cancer Screening
- Diabetes – Eye Exam
- Diabetes – HbA1c Screening
- Diabetes – Foot Exam
- Diabetes – Nephropathy Screening
- Chlamydia Screening (Peds, Med/Peds, and Family Medicine Only)
- HPV Vaccination (Peds, Med/Peds, and Family Medicine Only)

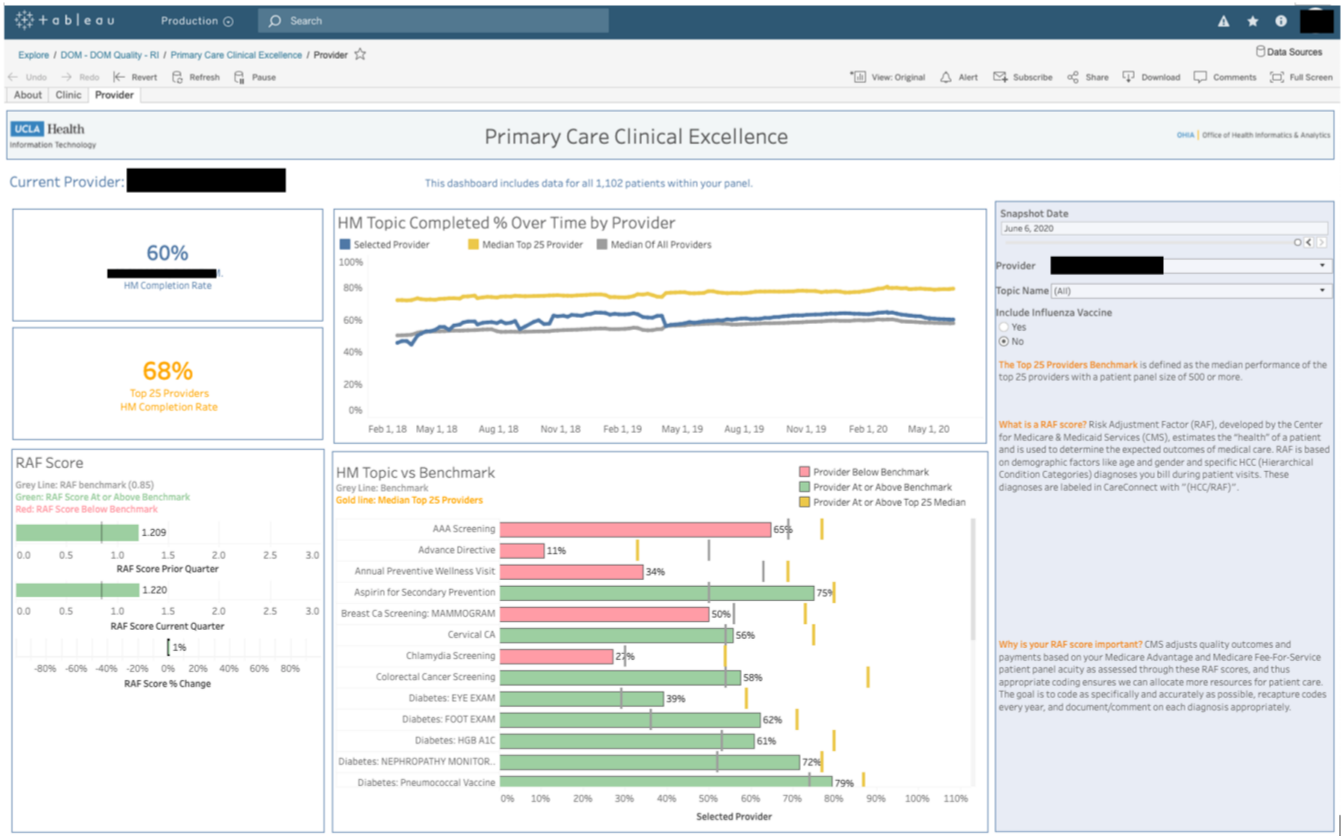
Please note, we recognize that the PCCE incentive program may not address important subtleties to medical management including engaging patients in shared decision making and risk-benefit discussions.

3. How do I track my overall performance?

To support your ongoing efforts, the Department will send you an email each month, giving you your health maintenance completion rate and providing best practice tips. This monthly update will also highlight the quality measure you are performing best on (relative to the median benchmark for that particular measure) as well as the two quality measures on which you could improve the most (relative to the median benchmark for these measures). These data are based on patients seen in the past three months.

Image of Dashboard

Below is an example image of the PCCE Dashboard (©Tableau) that physicians saw when they clicked on the dashboard link from their monthly email. The Dashboard breaks down how a physician is performing in each respective quality measure. Table A2-3 below shows the number of emails and dashboard clicks on a month-to-month basis.



Current Provider: [redacted]

This dashboard includes data for all 1,102 patients within your panel.

Snapshot Date: June 6, 2020

Provider: [redacted]

Topic Name: (All)

Include Influenza Vaccine

○ Yes

● No

The Top 25 Providers Benchmark is defined as the median performance of the top 25 providers with a patient panel size of 500 or more.

What is a RAF score? Risk Adjustment Factor (RAF), developed by the Center for Medicare & Medicaid Services (CMS), estimates the "health" of a patient and is used to determine the expected outcomes of medical care. RAF is based on demographic factors like age and gender and specific HCC (Hierarchical Condition Categories) diagnoses you bill during patient visits. These diagnoses are labeled in CareConnect with "(HCC/RAF)".

Why is your RAF score important? CMS adjusts quality outcomes and payments based on your Medicare Advantage and Medicare Fee-For-Service patient panel acuity as assessed through these RAF scores, and thus appropriate coding ensures we can allocate more resources for patient care. The goal is to code as specifically and accurately as possible, recapture codes every year, and document/comment on each diagnosis appropriately.

Table A2-3. Engagement with Intervention Emails

Month	Percentage of Opened Emails Across all Participants	Percentage of Dashboard Links Clicked Across all Participants
November 2019	73.4%	13.6%
December 2019	79.4%	14.1%
January 2020	71.4%	16.1%
February 2020	69.8%	15.1%
March 2020	68.8%	6.5%

Example of Monthly Best Practices Document

The monthly best practices documents were disseminated through the monthly performance emails as a link. These best practices included tips from UCLA physician champions on how to streamline certain orders and improve team-based primary care to ultimately improve HM completion rates. Displayed below is a page taken from the January 2020 Best Practices document.

1 Use the Annual Wellness Visit (AWV) to identify and close care gaps!

- Huddle with the MA or RN before each clinic session and identify patients coming in for the AWV, then review what care gaps need to be closed and divide and conquer!
- Use the Care Gaps SmartSet to quickly close care gaps.

- For a step-by-step guide, please see the Dec. 2019 Best Practices, Tip #2.

- Encourage your patients to sign up for MyChart and empower them to ask questions about their health. Ask the MA to help sign them up before they finish the visit!
- Use the dot-phrase [.HMLISTPT] in your note and After Visit Summary for a personalized patient-friendly list of health maintenance items and preventive care gaps.

Health Maintenance

Topic	Date Due
• Shingles Vaccine (1 of 2)	10/30/2003
• Pneumonia (PCV13 and/or PPSV23) Vaccine (2 of 2 - PPSV23)	02/08/2020
• Annual Preventive Wellness Visit	02/08/2020
• Colon Cancer Screening: Colonoscopy	03/08/2024
• Tdap/Td Vaccine (2 - Td)	09/10/2024
• Flu Vaccine	Completed
• Routine Hepatitis C Screening	Completed
• Abdominal Aortic Aneurysm Screening	Completed
• Statin prescribed for ASCVD Prevention or Treatment	Completed

Best Practice Tip

Ask patients to sign up for MyChart...[as it] is a good double check because it means a patient will bring up something on their record if it's incorrect. It empowers them to take ownership of their health.

– [redacted]

6. Leadership Support Training Intervention Details

Overview

In addition to the standard communication and peer comparison interventions, the physician leads and non-clinical managers in Condition 3 also received leadership training. Note that there were only 11 clinic leads in Condition 3 (and 33 clinic leads in the experiment) because non-clinical managers covered for physician leads in some clinics and some physician leads did not meet the experiment's inclusion criteria (e.g., due to FTE < 50%).

The aim of the leadership training intervention was to provide physician leads and non-clinical managers with the skills needed to foster a collaborative environment at their workplace, improve team-based primary care at their clinic, support their fellow PCPs, and engage their colleagues in a continuous cycle of quality improvement. The workshop curriculum guided them to formulate quality improvement goals for their clinic, design strategies to reach these goals, and disseminate best practices and key takeaways to the other PCPs at their clinic (e.g., core principles of team-based primary care, meaningful use of data to drive quality improvement).

As part of the leadership training intervention, physician leads and non-clinical managers within clinics randomized into Condition 3 attended two workshops on leadership and quality improvement. The two seminars occurred on December 3rd, 2019, and March 10th, 2020, respectively. Following the first workshop, physician leads and non-clinical managers also received additional one-on-one advice (via telephone calls, emails, and in-person meetings) from the DOM Quality Team. These meetings were intended to allow the clinic leadership team to revisit the takeaways from the workshop so they could formulate quality improvement goals and implementation plans to further improve team-based primary care at their clinics. All dyads were encouraged to schedule monthly all-clinic staff meetings to foster a communicative, positive team environment, discuss care gaps, and find strategies to enhance primary care quality.

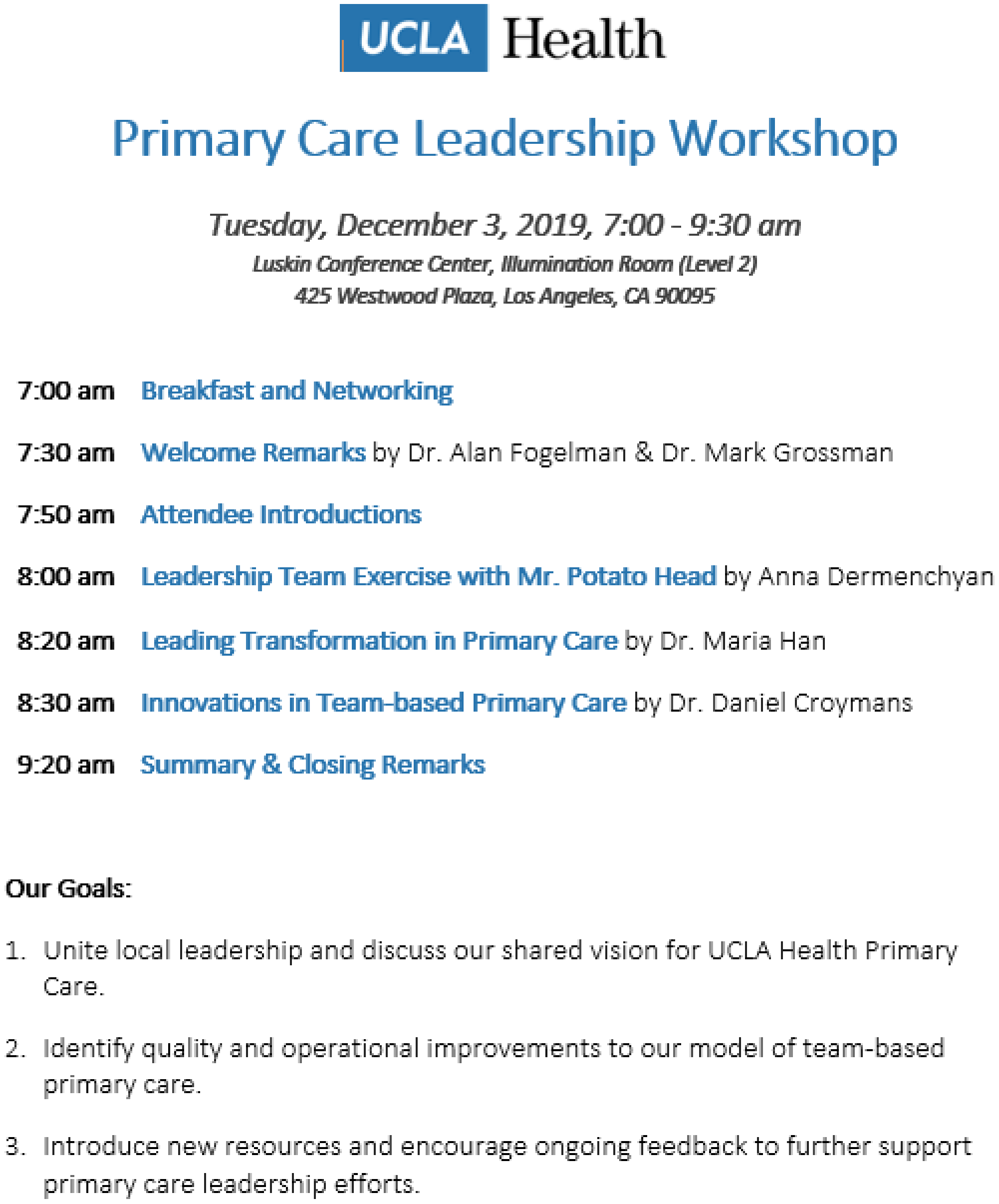
December 2019 Workshop

The first primary care leadership workshop was designed to help clinic physician leads and non-clinical managers recognize the importance of team-based primary care and encourage them to subsequently collaborate with clinical staff (e.g., front desk staff, nursing staff, other physicians) to more effectively foster team-based primary care within their own clinic. A copy of the workshop agenda can be found below.

- The workshop began by providing attendees a history and background on UCLA DOM's primary care network along with a discussion of the increasingly complex nature of primary care in recent years.
- Participants were then asked to participate in a Plan-Do-Study-Act (PDSA) team-building exercise with those at their table. This exercise (Constructing a Mr. Potato Head) is a quality improvement exercise for team building designed by the Institute for Healthcare Improvement.
    - Mr. Potato Head activity (see photo further below): Each team was tasked to try and construct the Mr. Potato Head as quickly as possible. After each timed attempt at constructing Mr. Potato Head, teams were encouraged to debrief with one another to identify what went right and what could be improved before beginning another attempt. Once all teams had made three attempts, the workshop attendees reconvened to debrief each other about the exercise. This was done to highlight the importance of communication and teamwork in complex tasks such as primary care.
- Next, the DOM Quality Team discussed the fundamental tenets of effective team-based primary care. These core tenets include:
    - Defined Purpose
    - Shared Goals
    - Clear Roles
    - Mutual Trust
    - Effective Communication
    - Measurable Processes and Outcomes
- Attendees were then split into groups for a breakout session where they brainstormed how to improve team-based primary care at UCLA Health using these core tenets. Following the brainstorming period, attendees were asked to report their suggestions to the larger group.
- To conclude the session, attendees were told to anticipate in one-on-one Quality Improvement (QI) meetings with the DOM Quality Team in the upcoming months. Additionally, attendees were asked to take 10-15 minutes out of their monthly clinic meetings to have data-driven conversations with their clinic staff. Whether or not these conversations occurred was not formally tracked.

Agenda for December 2019 Leadership Training Workshop

Displayed below is the agenda from the first Primary Care Leadership Workshop which took place in December 2019. Attendees (Condition 3 clinic physician leads and non-clinical managers) learned about UCLA's vision for primary care excellence, participated in team-building exercises, and brainstormed how to utilize best practices from the workshop in order to improve primary care practices at their respective clinics.



UCLA Health

Primary Care Leadership Workshop

Tuesday, December 3, 2019, 7:00 - 9:30 am
Luskin Conference Center, Illumination Room (Level 2)
425 Westwood Plaza, Los Angeles, CA 90095

7:00 am **Breakfast and Networking**

7:30 am **Welcome Remarks** by Dr. Alan Fogelman & Dr. Mark Grossman

7:50 am **Attendee Introductions**

8:00 am **Leadership Team Exercise with Mr. Potato Head** by Anna Dermenchyan

8:20 am **Leading Transformation in Primary Care** by Dr. Maria Han

8:30 am **Innovations in Team-based Primary Care** by Dr. Daniel Croymans

9:20 am **Summary & Closing Remarks**

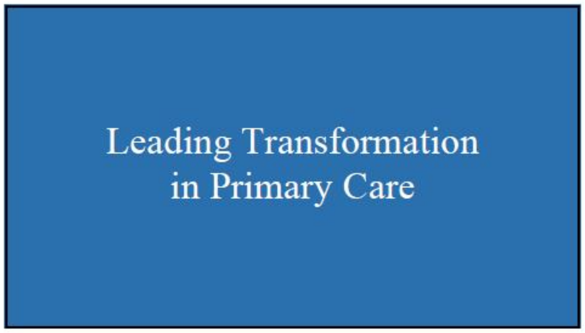
Our Goals:

1. Unite local leadership and discuss our shared vision for UCLA Health Primary Care.
2. Identify quality and operational improvements to our model of team-based primary care.
3. Introduce new resources and encourage ongoing feedback to further support primary care leadership efforts.

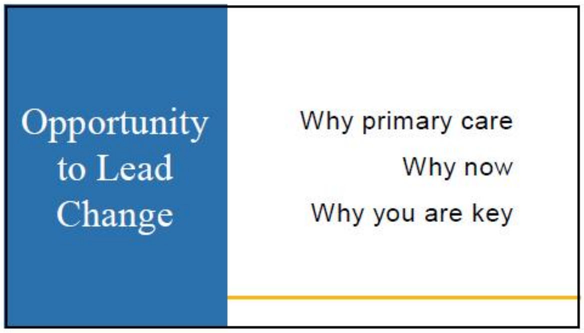
Mr. Potato Head Teamwork Exercise

As part of the Primary Care Leadership Workshop, attendees participated in a Mr. Potato head exercise in order to learn about Plan-Do-Study-Act cycles while emphasizing the importance of communication. Participants were informed that these skills could be used to experiment with solutions to clinic workflow challenges.

Presentation Slides from December 2019 Leadership Training Workshop



Leading Transformation in Primary Care

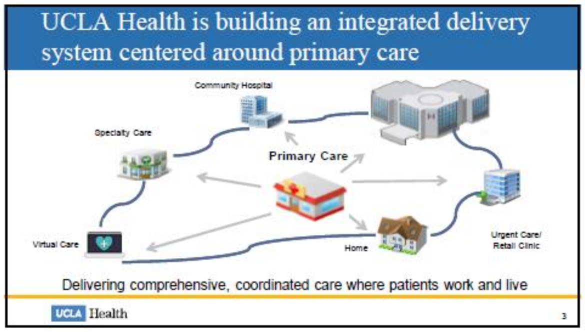


Opportunity to Lead Change

Why primary care

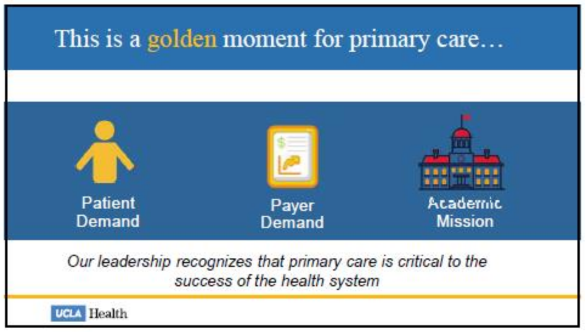
Why now

Why you are key



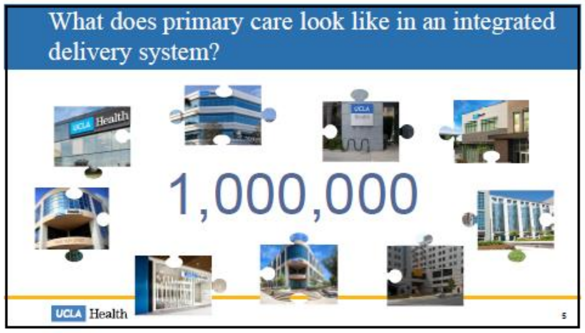
UCLA Health is building an integrated delivery system centered around primary care

Delivering comprehensive, coordinated care where patients work and live



This is a golden moment for primary care…

Our leadership recognizes that primary care is critical to the success of the health system



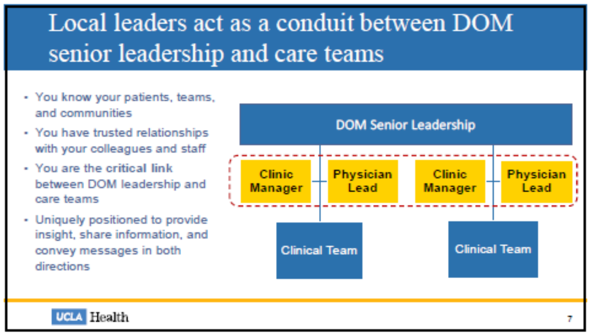
What does primary care look like in an integrated delivery system?

1,000,000

Having strong local leaders in our communities is paramount to our success

UCLA Health Expansion, 2015 to 2018

Central leaders can't begin to have the impact in your clinics that you can



Local leaders act as a conduit between DOM senior leadership and care teams

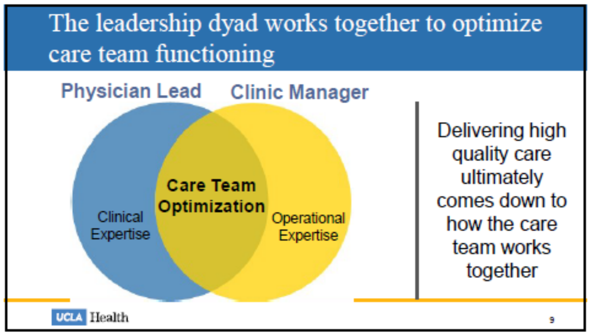
- You know your patients, teams, and communities
- You have trusted relationships with your colleagues and staff
- You are the **critical link** between DOM leadership and care teams
- Uniquely positioned to provide insight, share information, and convey messages in both directions

7

Local clinical-operational leadership dyads support integrated care delivery

- Equal partnership between a clinical leader and an operational leader
- Complementary skills and strengths
- A highly functioning dyad makes decisions, sets priorities, and holds the team accountable together
- **Ideally brings together the best in clinical thinking with the best in operational thinking for the best in patient care**

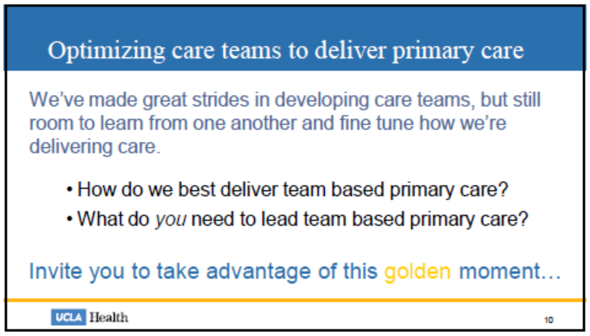
8



The leadership dyad works together to optimize care team functioning

Delivering high quality care ultimately comes down to how the care team works together

9



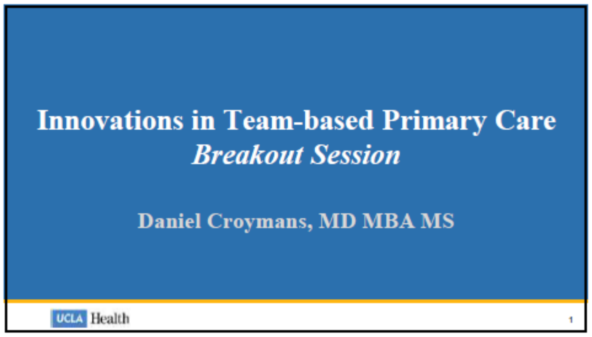
Optimizing care teams to deliver primary care

We've made great strides in developing care teams, but still room to learn from one another and fine tune how we're delivering care.

- How do we best deliver team based primary care?
- What do *you* need to lead team based primary care?

Invite you to take advantage of this golden moment…

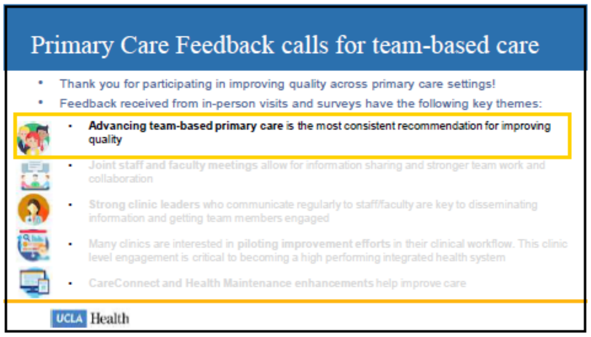
10



Innovations in Team-based Primary Care
Breakout Session

Daniel Croymans, MD MBA MS

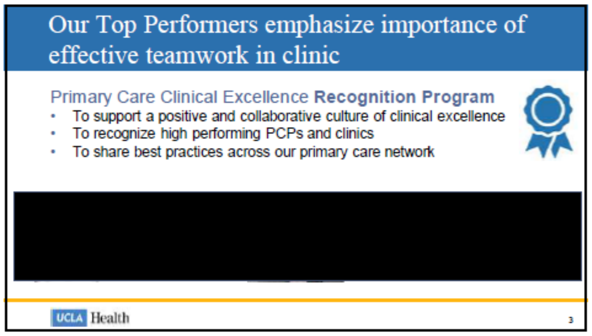
1



Primary Care Feedback calls for team-based care

- Thank you for participating in improving quality across primary care settings!
- Feedback received from in-person visits and surveys have the following key themes:
  - **Advancing team-based primary care** is the most consistent recommendation for improving quality
  - Joint staff and faculty meetings allow for information sharing and stronger team work and collaboration
  - Strong clinic leaders who communicate regularly to staff/faculty are key to disseminating information and getting team members engaged
  - Many clinics are interested in piloting improvement efforts in their clinical workflow. This clinic level engagement is critical to becoming a high performing integrated health system
  - CareConnect and Health Maintenance enhancements help improve care

2

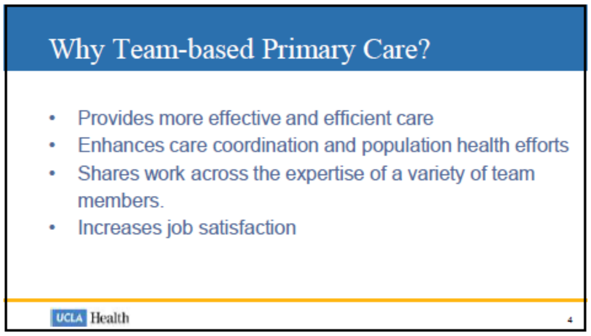


Our Top Performers emphasize importance of effective teamwork in clinic

Primary Care Clinical Excellence **Recognition Program**
- To support a positive and collaborative culture of clinical excellence
- To recognize high performing PCPs and clinics
- To share best practices across our primary care network

[redacted]

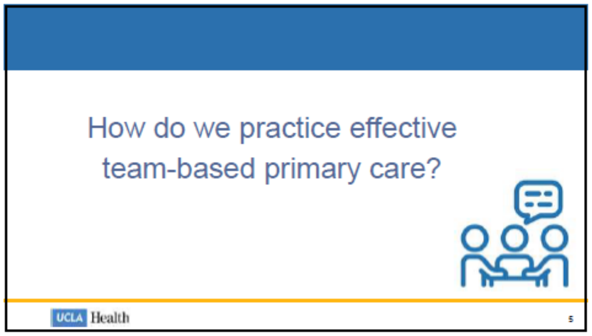
3



Why Team-based Primary Care?

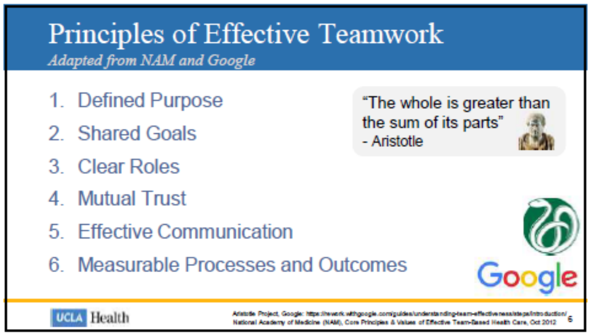
- Provides more effective and efficient care
- Enhances care coordination and population health efforts
- Shares work across the expertise of a variety of team members.
- Increases job satisfaction

4



How do we practice effective team-based primary care?

5



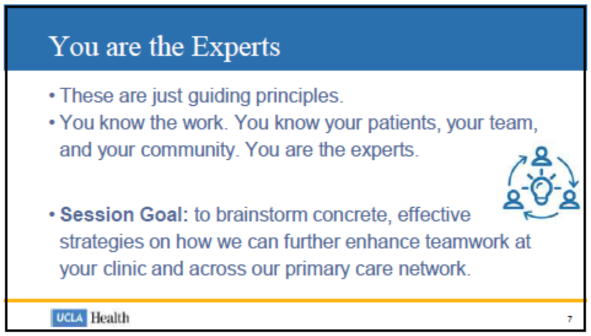
Principles of Effective Teamwork
Adapted from NAM and Google

1. Defined Purpose
2. Shared Goals
3. Clear Roles
4. Mutual Trust
5. Effective Communication
6. Measurable Processes and Outcomes

"The whole is greater than the sum of its parts"
- Aristotle

Aristotle Project, Google: https://rework.withgoogle.com/guides/understanding-team-effectiveness/steps/introduction/
National Academy of Medicine (NAM), Core Principles & Values of Effective Team-Based Health Care, Oct 2012

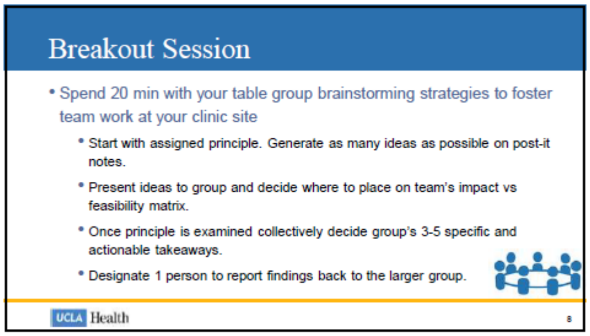
6



You are the Experts

- These are just guiding principles.
- You know the work. You know your patients, your team, and your community. You are the experts.
- **Session Goal:** to brainstorm concrete, effective strategies on how we can further enhance teamwork at your clinic and across our primary care network.

UCLA Health



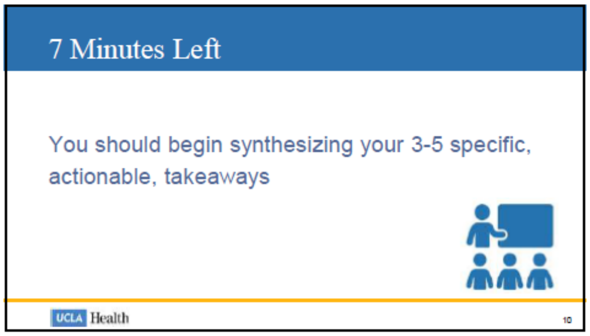
Breakout Session

- Spend 20 min with your table group brainstorming strategies to foster team work at your clinic site
  - Start with assigned principle. Generate as many ideas as possible on post-it notes.
  - Present ideas to group and decide where to place on team's impact vs feasibility matrix.
  - Once principle is examined collectively decide group's 3-5 specific and actionable takeaways.
  - Designate 1 person to report findings back to the larger group.

UCLA Health

Begin!

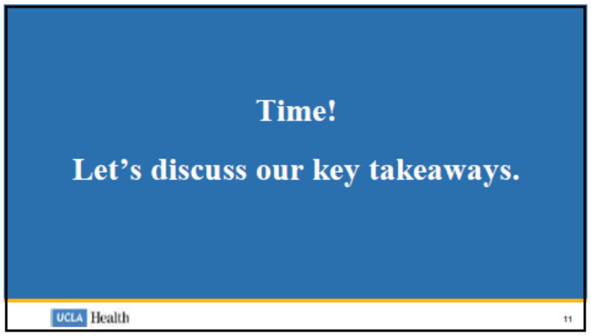
UCLA Health



7 Minutes Left

You should begin synthesizing your 3-5 specific, actionable, takeaways

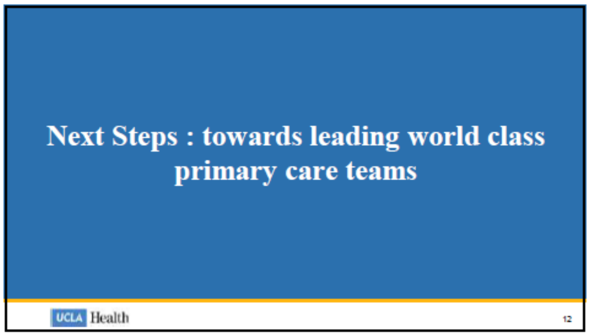
UCLA Health



Time!

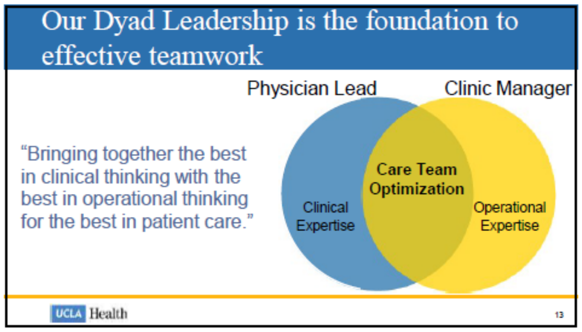
Let's discuss our key takeaways.

UCLA Health



Next Steps : towards leading world class primary care teams

UCLA Health

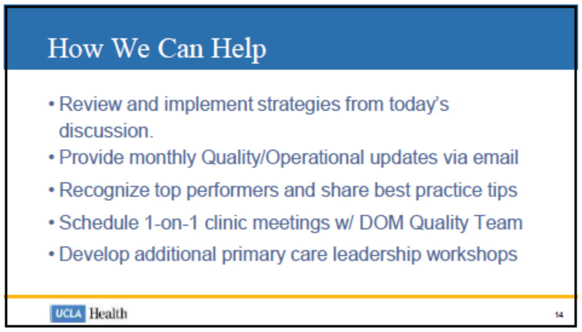


Our Dyad Leadership is the foundation to effective teamwork

"Bringing together the best in clinical thinking with the best in operational thinking for the best in patient care."

UCLA Health 13

13

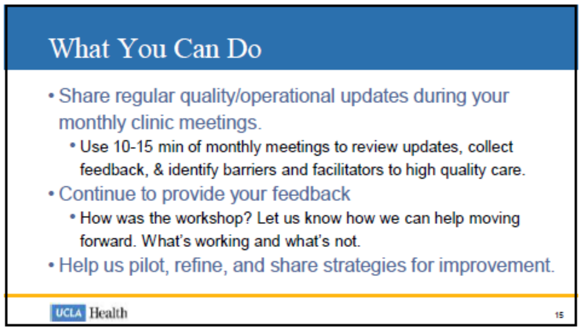


How We Can Help

- Review and implement strategies from today's discussion.
- Provide monthly Quality/Operational updates via email
- Recognize top performers and share best practice tips
- Schedule 1-on-1 clinic meetings w/ DOM Quality Team
- Develop additional primary care leadership workshops

UCLA Health 14

14

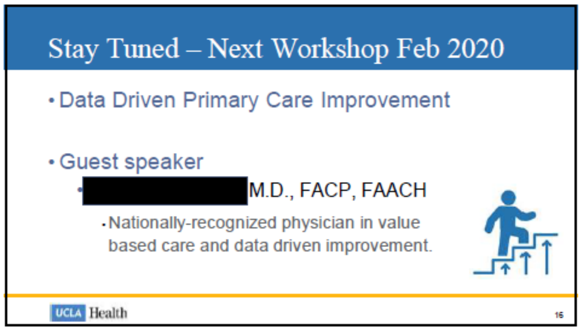


What You Can Do

- Share regular quality/operational updates during your monthly clinic meetings.
  - Use 10-15 min of monthly meetings to review updates, collect feedback, & identify barriers and facilitators to high quality care.
- Continue to provide your feedback
  - How was the workshop? Let us know how we can help moving forward. What's working and what's not.
- Help us pilot, refine, and share strategies for improvement.

UCLA Health 15

15



Stay Tuned – Next Workshop Feb 2020

- Data Driven Primary Care Improvement
- Guest speaker
  - [redacted] M.D., FACP, FAACH
    - Nationally-recognized physician in value based care and data driven improvement.

UCLA Health 16

16

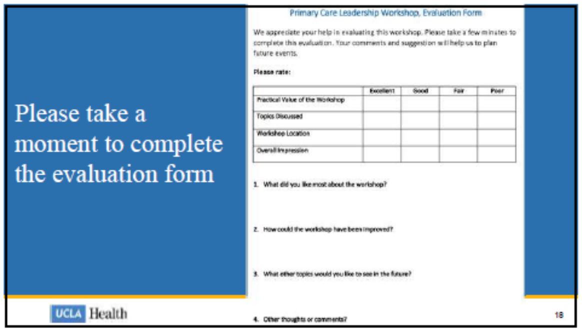
Thank you for your time and leadership!

We are dedicated to working with you to build the best system of care for our patients, providers, and staff.

DOMQuality@mednet.ucla.edu

UCLA Health 17

17



Please take a moment to complete the evaluation form

Primary Care Leadership Workshop, Evaluation Form

We appreciate your help in evaluating this workshop. Please take a few minutes to complete this evaluation. Your comments and suggestion will help us to plan future events.

Please rate:

	Excellent	Good	Fair	Poor
Practical Value of the Workshop				
Topics Discussed				
Workshop Location				
Overall Impression				

1. What did you like most about the workshop?
2. How could the workshop have been improved?
3. What other topics would you like to see in the future?
4. Other thoughts or comments?

UCLA Health 18

18

*Note the "Feb 2020" date listed on Slide 16 was tentative. Following the first workshop in December 2019, DOM leadership decided to have the second workshop take place in March 2020.

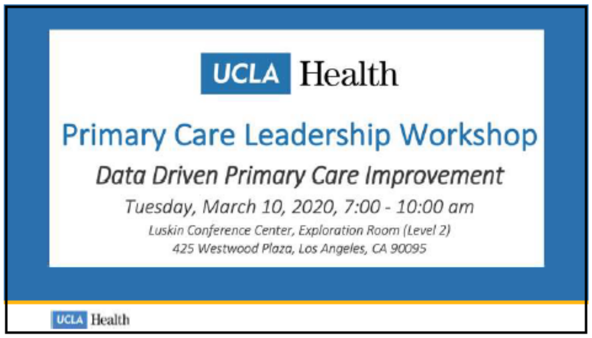
March 2020 Leadership Training Workshop

The second primary care leadership workshop aimed to continue the conversation about fostering team-based primary care and translating these efforts into practice so as to improve clinical quality measures at a clinic level. Attendees were encouraged to take initiative to improve clinic performance in ways they saw fit for their clinic.

- This second primary care leadership workshop included a guest speaker who is a widely recognized expert in designing and implementing pay-for-performance models in primary care. The central message of this part of the workshop was that successful primary care networks would foster the following:
    - Core values
    - Team-based care
    - Senior management and board buy-in
    - A non-judgmental workspace
- Next, a UCLA DOM member shared an experience about how they track up-to-date information on their patients' health statuses.
- Attendees were then split up for a breakout session. Each table was assigned a clinical quality measure (HbA1c screening, BP control, etc.) and asked to identify current primary care gaps and craft solutions to address them. After brainstorming ideas, each group then reported their findings to the larger group.
- Finally, attendees were reminded to utilize available data and their own clinical experience to identify best practices for their own clinics in order to deliver high quality primary care to their patients.

Attendees were trained on how to guide conversations with their co-workers whereby they could formulate performance/quality improvement goals, design effective strategies to reach these goals, and track their clinic's progress.

Presentation Slides from March 2020 Leadership Training Workshop



UCLA Health

Primary Care Leadership Workshop

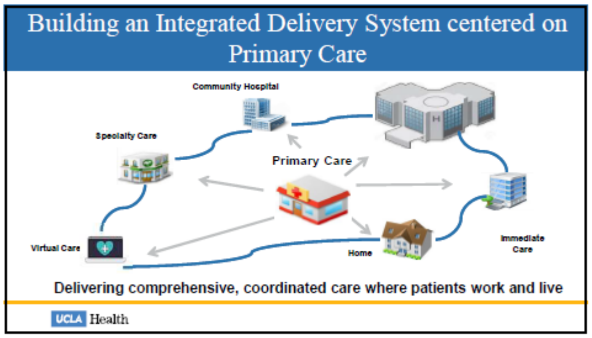
Data Driven Primary Care Improvement

Tuesday, March 10, 2020, 7:00 - 10:00 am

Luskin Conference Center, Exploration Room (Level 2)
425 Westwood Plaza, Los Angeles, CA 90095

UCLA Health

1

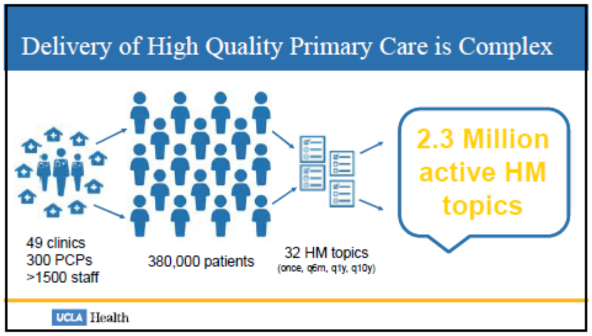


Building an Integrated Delivery System centered on Primary Care

Delivering comprehensive, coordinated care where patients work and live

UCLA Health

2



Delivery of High Quality Primary Care is Complex

49 clinics
300 PCPs
>1500 staff

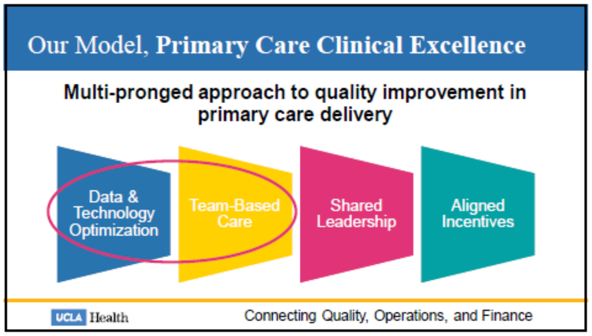
380,000 patients

32 HM topics
(once, q6m, q1y, q10y)

2.3 Million active HM topics

UCLA Health

3



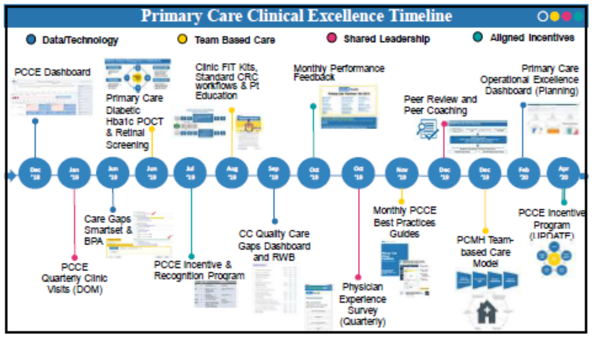
Our Model, **Primary Care Clinical Excellence**

Multi-pronged approach to quality improvement in primary care delivery

UCLA Health

Connecting Quality, Operations, and Finance

4

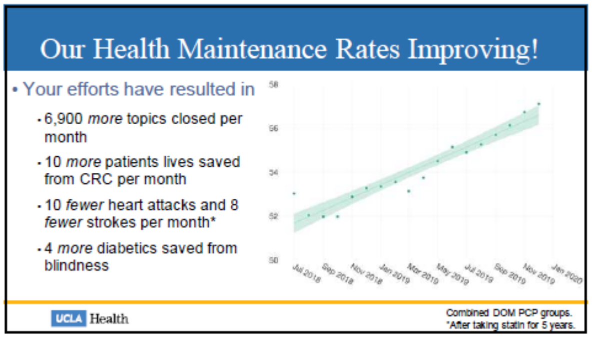


Primary Care Clinical Excellence Timeline

● Data/Technology
● Team Based Care
● Shared Leadership
● Aligned Incentives

- PCCE Dashboard
- PCCE Quarterly Clinic Visits (DOM)
- Care Gaps Smartset & BPA
- Primary Care Diabetic Hba1c POCT & Retinal Screening
- PCCE Incentive & Recognition Program
- Clinic FIT Kits, Standard CRC workflows & Pt Education
- CC Quality Care Gaps Dashboard and RWB
- Monthly Performance Feedback
- Physician Experience Survey (Quarterly)
- Monthly PCCE Best Practices Guides
- Peer Review and Peer Coaching
- PCMH Team-based Care Model
- Primary Care Operational Excellence Dashboard (Planning)
- PCCE Incentive Program (UPDATE)

5



Our Health Maintenance Rates Improving!

- Your efforts have resulted in
  - 6,900 *more* topics closed per month
  - 10 *more* patients lives saved from CRC per month
  - 10 *fewer* heart attacks and 8 *fewer* strokes per month*
  - 4 *more* diabetics saved from blindness

UCLA Health

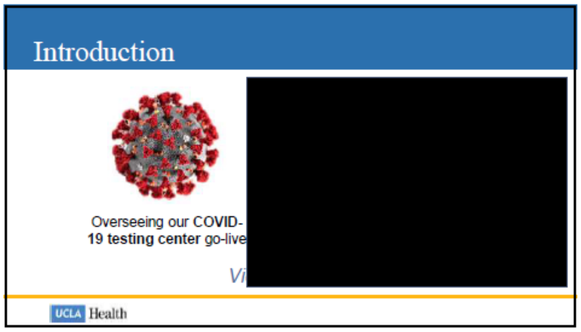
Combined DOM PCP groups.
*After taking statin for 5 years.

6

Today's Agenda

7:00 am	Breakfast and Networking
7:30 am	Welcome Remarks by Drs. Daniel Croymans and Mark Grossman
7:40 am	Attendee Introductions
7:50 am	Data-Driven Primary Care Improvement by Dr. Howard Beckman
8:30 am	Break
8:45 am	Innovations in Diabetes Care Management by Dr. Ben Waterman
9:00 am	Breakout Session by Jeff Butler, Anna Dermenchyan, & Drs. Howard Beckman, Daniel Croymans, Maria Han, and Ben Waterman
9:30 am	Report Out by Individual Groups
9:50 am	Summary & Closing Remarks by Dr. Maria Han

UCLA Health

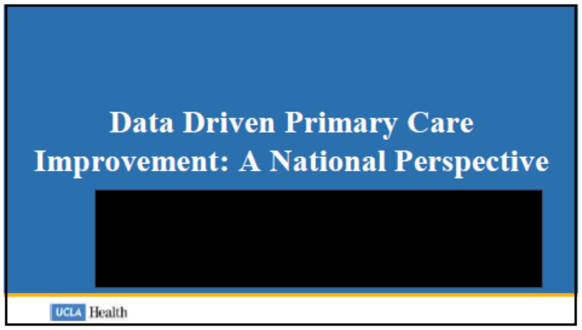


Introduction

Overseeing our COVID-19 testing center go-live

Vi [redacted]

UCLA Health

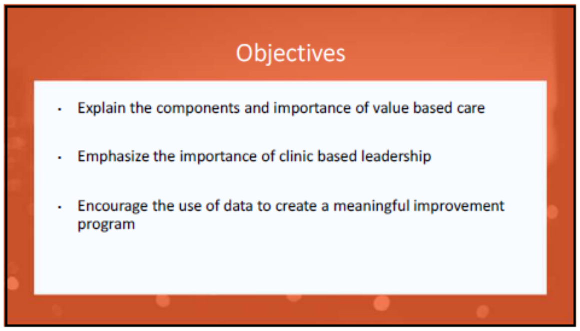


Data Driven Primary Care Improvement: A National Perspective

[redacted]

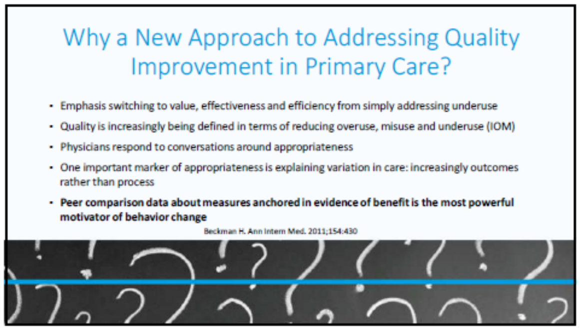
UCLA Health

The Current State of Affairs



Objectives

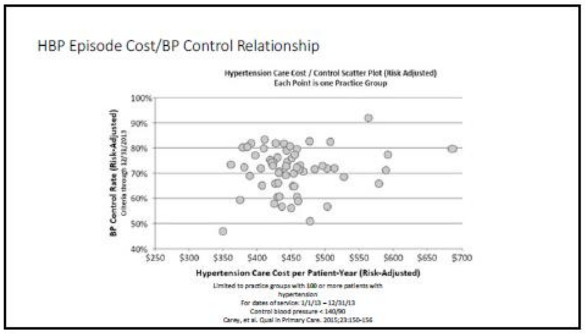
- Explain the components and importance of value based care
- Emphasize the importance of clinic based leadership
- Encourage the use of data to create a meaningful improvement program



Why a New Approach to Addressing Quality Improvement in Primary Care?

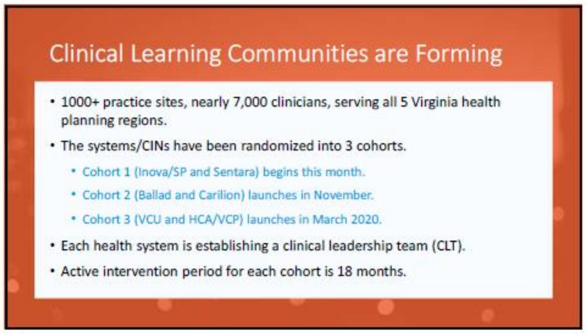
- Emphasis switching to value, effectiveness and efficiency from simply addressing underuse
- Quality is increasingly being defined in terms of reducing overuse, misuse and underuse (IOM)
- Physicians respond to conversations around appropriateness
- One important marker of appropriateness is explaining variation in care: increasingly outcomes rather than process
- **Peer comparison data about measures anchored in evidence of benefit is the most powerful motivator of behavior change**

Beckman H. Ann Intern Med. 2011;154:430



HBP Episode Cost/BP Control Relationship

Limited to practice groups with 100 or more patients with hypertension
For dates of service: 1/1/13 – 12/31/13
Control blood pressure < 140/90
Carey, et al. Qual in Primary Care. 2015;23:150-156

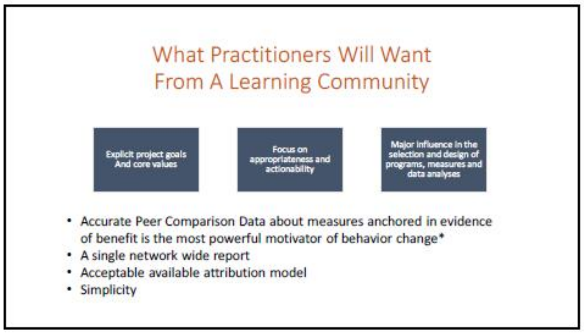


Clinical Learning Communities are Forming

- 1000+ practice sites, nearly 7,000 clinicians, serving all 5 Virginia health planning regions.
- The systems/CINs have been randomized into 3 cohorts.
  - Cohort 1 (Inova/SP and Sentara) begins this month.
  - Cohort 2 (Ballad and Carilion) launches in November.
  - Cohort 3 (VCU and HCA/VCP) launches in March 2020.
- Each health system is establishing a clinical leadership team (CLT).
- Active intervention period for each cohort is 18 months.

Clinical Learning Community

BalladHealth
CARILION CLINIC
INOVA
HCA
Hospital Corporation of America
SENTARA
VCUHealth
Sentara Quality Care Network
SIGNATURE
VIRGINIACARE PARTNERS



What Practitioners Will Want From A Learning Community

- Accurate Peer Comparison Data about measures anchored in evidence of benefit is the most powerful motivator of behavior change*
- A single network wide report
- Acceptable available attribution model
- Simplicity

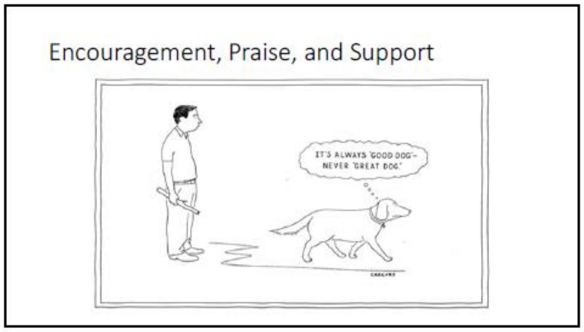
Getting Started: Leadership's Role

Getting Started: Senior Leadership's Role

- Publicize the project **and its value** through organizational outlets (avoid surprises and **normalize** the process):
  - Board meetings
  - Committee or Specialty meetings
  - Newsletter
  - Identify Project Champions
  - Praise successes

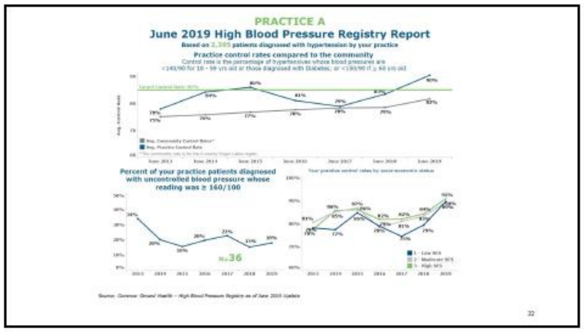
Getting Started: Leadership's Role (cont'd)

- Clarifying core values
- Providing needed resources
- Announcing the Project Leadership Team
- Avoiding triangulation
- Celebrating success



Encouragement, Praise, and Support

Does It Work?: Examples of Success



PRACTICE A

June 2019 High Blood Pressure Registry Report

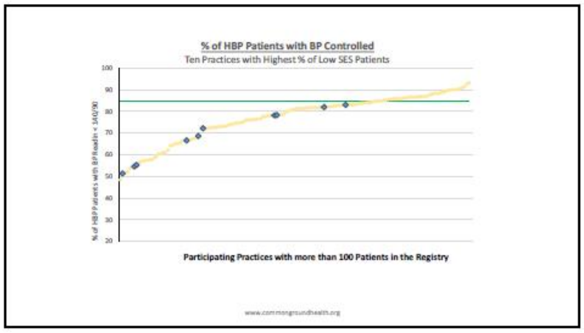
Based on 2,305 patients diagnosed with hypertension by your practice

Practice control rates compared to the community

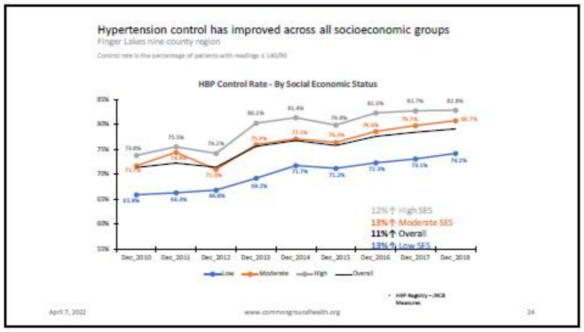
Control rate is the percentage of hypertensives whose blood pressures are <140/90 for 18 - 59 yrs old or those diagnosed with Diabetes; or <150/90 if ≥ 60 yrs old

Percent of your practice patients diagnosed with uncontrolled blood pressure whose reading was ≥ 160/100

Source: Common Ground Health – High Blood Pressure Registry as of June 2019 Update



www.commongroundhealth.org



Hypertension control has improved across all socioeconomic groups

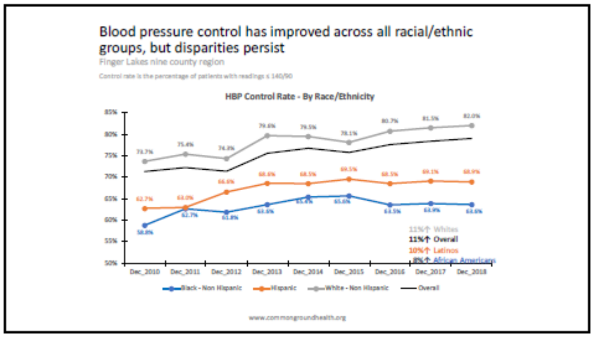
Finger Lakes nine county region

Control rate is the percentage of patients with readings ≤ 140/90

- HBP Registry – JNC8 Measures

April 7, 2022

www.commongroundhealth.org

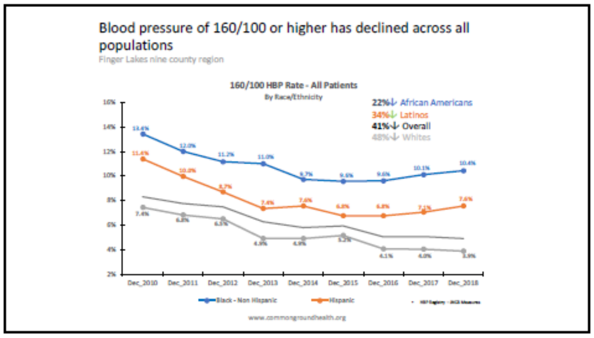


Blood pressure control has improved across all racial/ethnic groups, but disparities persist

Finger Lakes nine county region

Control rate is the percentage of patients with readings ≤ 140/90

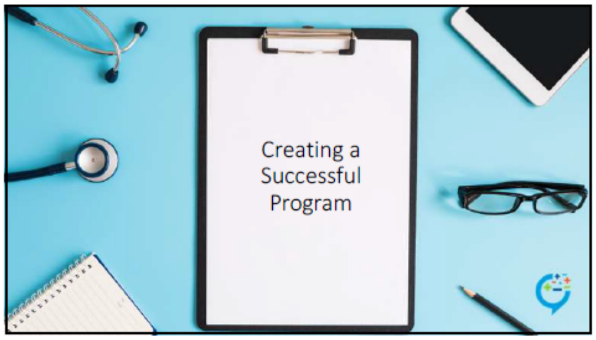
www.commongroundhealth.org



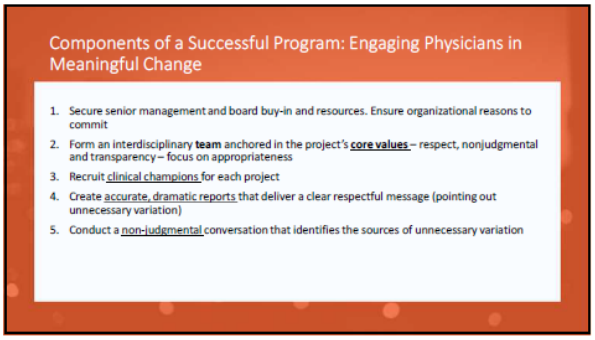
Blood pressure of 160/100 or higher has declined across all populations

Finger Lakes nine county region

www.commongroundhealth.org



Creating a Successful Program



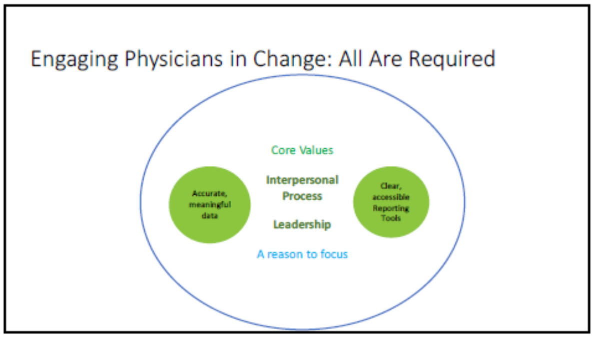
Components of a Successful Program: Engaging Physicians in Meaningful Change

1. Secure senior management and board buy-in and resources. Ensure organizational reasons to commit
2. Form an interdisciplinary **team** anchored in the project's **core values** – respect, nonjudgmental and transparency – focus on appropriateness
3. Recruit clinical champions for each project
4. Create accurate, dramatic reports that deliver a clear respectful message (pointing out unnecessary variation)
5. Conduct a non-judgmental conversation that identifies the sources of unnecessary variation

Components of a Successful Program: Engaging Physicians in Meaningful Change

Cont'd

6. Collaboratively construct a way to address and reduce that unnecessary variation (a quality improvement plan)
7. Offer physicians on-going feedback through a respectful process of sharing data and facilitating improvement
8. Communicate regularly with project team, physicians and key sponsors on program progress and outcomes
9. Use tracking tools to monitor and report interim measures of success, (reaching targets, % improvement, $ savings and ROI)
10. Praise success (ex. Newsletters, bonuses, plaques)!



Engaging Physicians in Change: All Are Required

A Reason to Focus

- Reducing overuse and underuse of services?
- Self satisfaction in delivering appropriate care?
- Public/Professional recognition for delivering appropriate care?
- Patient Safety?
- Financial incentives?
- Peer Recognition?
- Reducing the cost of care?
- **What will you promote in your practice?**

Beckman H. Lost in Translation, Ann Intern Med. 2011;154:430

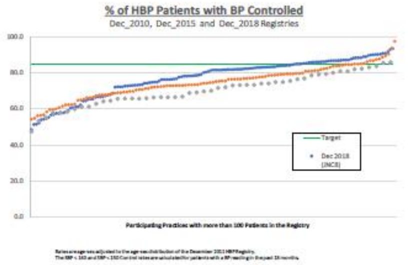
What Does Clinical Variation Look Like: At What Level to Report?

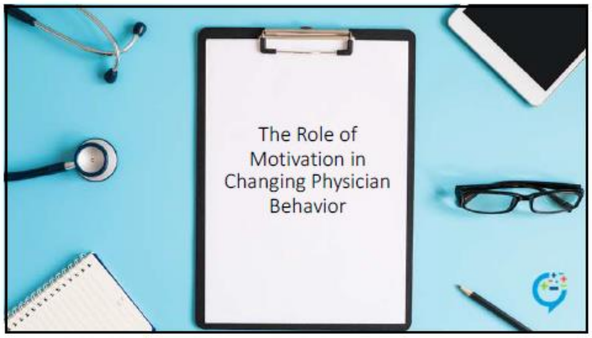
Using Variation Wisely

Why So Much Variation?

Basis of Decisions	Number of Decisions*	% of Total
Experience/Anecdote	441	37.1
Arbitrary/Instinct	175	14.7
Trained to do it	173	14.6
General Study	146	12.3
First Principles	146	12.3
Limited Study	61	5.1
Specific Study	34	2.9
Parental Preference	6	0.5
For Research	4	0.3
Avoid a Lawsuit	2	0.2
* Rounded to the nearest integer	1188	100.0

Darst JR, et al. Deciding without Data. Congenital Heart disease. 2010;5:339





The Role of Motivation in Changing Physician Behavior

Self Determination Theory

- Developed by Ed Deci, Ph.D. and Richard Ryan, Ph.D.
- Proposes that internal motivation trumps external motivation
- Central for working with team members and practitioners
- Defines three areas responsible for internal motivation
  - Competence
  - Autonomy
  - Relatedness
  - In the context of synchronous core values

Competence

- Asking someone to accomplish something they believe is possible
- The need to feel that one can reliably produce desired outcomes and/or avoid negative outcomes

Autonomy

- Being given the chance to discover how to solve a problem; encouraged to own the solution
- Autonomy relates to the feeling that one is
  acting in accord with one's sense of self
- A sense of choosing rather than feeling compelled or controlled

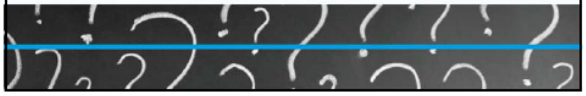
Relatedness

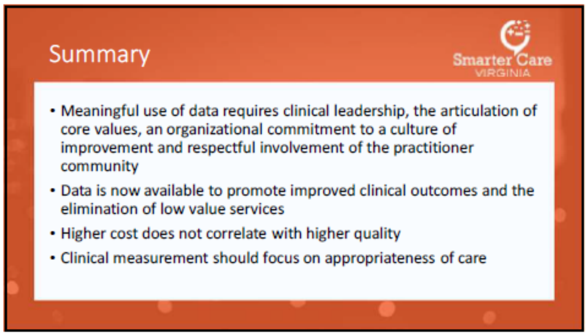
- Believing one is being asked to be part of a larger task, goal, community (Doing meaningful work)
- Context values – Believing in the team asking for the effort. Feeling that the community involved in the project shares reasons for participating and conducts its work responsibly

Self-Determination and Motivation

- https://www.youtube.com/watch?v=u6XAPnuFjJc

Comments? Concerns? Relevance to UCLA Operation?

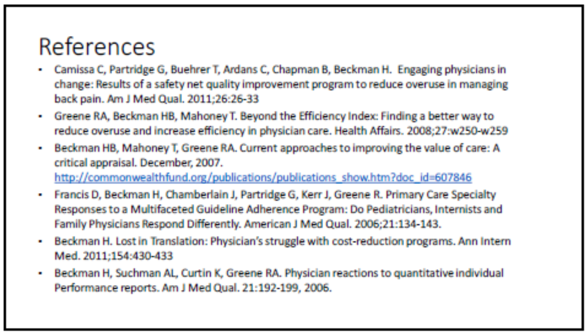




Summary

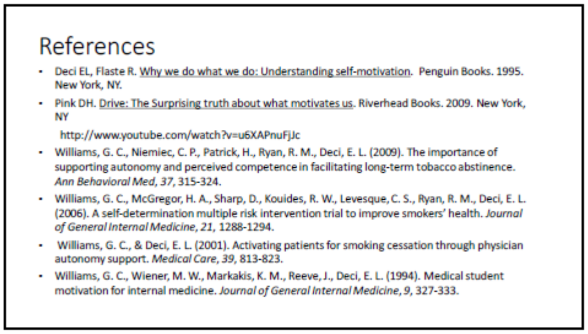
Smarter Care VIRGINIA

- Meaningful use of data requires clinical leadership, the articulation of core values, an organizational commitment to a culture of improvement and respectful involvement of the practitioner community
- Data is now available to promote improved clinical outcomes and the elimination of low value services
- Higher cost does not correlate with higher quality
- Clinical measurement should focus on appropriateness of care



References

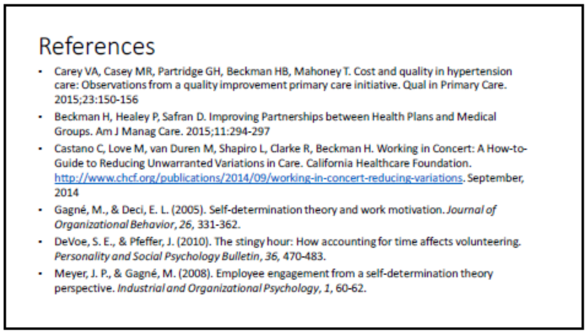
- Camissa C, Partridge G, Buehrer T, Ardans C, Chapman B, Beckman H. Engaging physicians in change: Results of a safety net quality improvement program to reduce overuse in managing back pain. Am J Med Qual. 2011;26:26-33
- Greene RA, Beckman HB, Mahoney T. Beyond the Efficiency Index: Finding a better way to reduce overuse and increase efficiency in physician care. Health Affairs. 2008;27:w250-w259
- Beckman HB, Mahoney T, Greene RA. Current approaches to improving the value of care: A critical appraisal. December, 2007. http://commonwealthfund.org/publications/publications_show.htm?doc_id=607846
- Francis D, Beckman H, Chamberlain J, Partridge G, Kerr J, Greene R. Primary Care Specialty Responses to a Multifaceted Guideline Adherence Program: Do Pediatricians, Internists and Family Physicians Respond Differently. American J Med Qual. 2006;21:134-143.
- Beckman H. Lost in Translation: Physician's struggle with cost-reduction programs. Ann Intern Med. 2011;154:430-433
- Beckman H, Suchman AL, Curtin K, Greene RA. Physician reactions to quantitative individual Performance reports. Am J Med Qual. 21:192-199, 2006.



References

- Deci EL, Flaste R. Why we do what we do: Understanding self-motivation. Penguin Books. 1995. New York, NY.
- Pink DH. Drive: The Surprising truth about what motivates us. Riverhead Books. 2009. New York, NY

  http://www.youtube.com/watch?v=u6XAPnuFjJc
- Williams, G. C., Niemiec, C. P., Patrick, H., Ryan, R. M., Deci, E. L. (2009). The importance of supporting autonomy and perceived competence in facilitating long-term tobacco abstinence. *Ann Behavioral Med*, *37*, 315-324.
- Williams, G. C., McGregor, H. A., Sharp, D., Kouides, R. W., Levesque, C. S., Ryan, R. M., Deci, E. L. (2006). A self-determination multiple risk intervention trial to improve smokers' health. *Journal of General Internal Medicine*, *21*, 1288-1294.
- Williams, G. C., & Deci, E. L. (2001). Activating patients for smoking cessation through physician autonomy support. *Medical Care*, *39*, 813-823.
- Williams, G. C., Wiener, M. W., Markakis, K. M., Reeve, J., Deci, E. L. (1994). Medical student motivation for internal medicine. *Journal of General Internal Medicine*, *9*, 327-333.



References

- Carey VA, Casey MR, Partridge GH, Beckman HB, Mahoney T. Cost and quality in hypertension care: Observations from a quality improvement primary care initiative. Qual in Primary Care. 2015;23:150-156
- Beckman H, Healey P, Safran D. Improving Partnerships between Health Plans and Medical Groups. Am J Manag Care. 2015;11:294-297
- Castano C, Love M, van Duren M, Shapiro L, Clarke R, Beckman H. Working in Concert: A How-to-Guide to Reducing Unwarranted Variations in Care. California Healthcare Foundation. http://www.chcf.org/publications/2014/09/working-in-concert-reducing-variations. September, 2014
- Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior*, *26*, 331-362.
- DeVoe, S. E., & Pfeffer, J. (2010). The stingy hour: How accounting for time affects volunteering. *Personality and Social Psychology Bulletin*, *36*, 470-483.
- Meyer, J. P., & Gagné, M. (2008). Employee engagement from a self-determination theory perspective. *Industrial and Organizational Psychology*, *1*, 60-62.

7. Quarterly Survey Materials, Schedule, and Completion Rates

Quarterly Survey Questions

The questions from the quarterly survey are displayed below (excluding the questions that measured physicians' self-reported attendance at professional activities and committees). See questions 8, 15, and 16 for our measures of perceived leadership support, job satisfaction, and burnout, respectively. Note that CareConnect is the name of UCLA Health's Electronic Health Record system (Epic Systems, ©1979).

1. The degree to which my care team works efficiently together is:
    - Poor
    - Marginal
    - Satisfactory
    - Good
    - Optimal

2. My proficiency with using CareConnect is:
    - Poor
    - Marginal
    - Satisfactory
    - Good
    - Optimal

3. I have frequent opportunities to make improvements at my clinic.
    - Strongly Disagree
    - Disagree
    - Neither Agree Nor Disagree
    - Agree
    - Strongly Agree

4. I am involved in deciding on changes that affect my work and care team.
    - Strongly Disagree
    - Disagree
    - Neither Agree Nor Disagree
    - Agree
    - Strongly Agree

5. I have adequate performance feedback and best practice guidelines to help me provide high quality care.
    - Strongly Disagree
    - Disagree
    - Neither Agree Nor Disagree
    - Agree
    - Strongly Agree

6. I am confident in my ability to use performance feedback and best practice guidelines to help me provide high quality care.

- Strongly Disagree
- Disagree
- Neither Agree Nor Disagree
- Agree
- Strongly Agree

7. I feel supported, understood, and valued by my *work colleagues*.

- Strongly Disagree
- Disagree
- Neither Agree Nor Disagree
- Agree
- Strongly Agree

8. I feel supported, understood, and valued by my *department leaders*.

- Strongly Disagree
- Disagree
- Neither Agree Nor Disagree
- Agree
- Strongly Agree

Most people compare themselves from time to time with others. For example, they may compare the way they feel, their opinions, their abilities, and/or their situation with those of other people. There is nothing particularly 'good' or 'bad' about this type of comparison, and some people do it more than others.

We would like to find out how often you compare yourself with other people. To do that we would like to ask you to indicate how much you agree with each statement below.

9. If I want to find out how well I have done something, I compare what I have done with how others have done.

- Strongly Disagree
- Disagree
- Neither Agree Nor Disagree
- Agree
- Strongly Agree

10. If I want to learn more about something, I try to find out what others think about it.

- Strongly Disagree
- Disagree
- Neither Agree Nor Disagree
- Agree
- Strongly Agree

Please rate the extent to which each reason below describes why you are currently engaged in your profession:

11. Because I enjoy this work very much.
    - Not at all
    - Very Little
    - A Little
    - Moderately
    - Strongly
    - Very Strongly
    - Exactly

12. Because this job fits my personal values.
    - Not at all
    - Very Little
    - A Little
    - Moderately
    - Strongly
    - Very Strongly
    - Exactly

13. Because this job affords me a desirable standard of living.
    - Not at all
    - Very Little
    - A Little
    - Moderately
    - Strongly
    - Very Strongly
    - Exactly

14. Because my reputation depends on it.
    - Not at all
    - Very Little
    - A Little
    - Moderately
    - Strongly
    - Very Strongly
    - Exactly

15. Taking everything into consideration, how do you feel about your job as a whole?
    - Extremely Dissatisfied
    - Dissatisfied
    - Somewhat Dissatisfied
    - Neutral
    - Somewhat Satisfied
    - Satisfied

- Extremely Satisfied

16. Overall, based on your definition of burnout, how would you rate your level of burnout?
    - I enjoy my work. I have no symptoms of burnout.
    - Occasionally I am under stress, and I don't always have as much energy as I once did, but I don't feel burned out.
    - I am definitely burning out and have one or more symptoms of burnout, such as physical and emotional exhaustion.
    - The symptoms of burnout that I am experiencing will not go away. I think about frustration at work a lot.
    - I feel completely burned out and often wonder if I can go on. I am at the point where I may need some changes or may need to seek some sort of help.

17. In the past three months, what were the two most significant barriers that hindered your delivery of excellent patient care?

    ____
    ____
    ____
    ____

Demographic Information

Please answer the following confidential demographic questions. You will only need to complete this section once. This section will help us identify how population characteristics might relate to physician experiences.

1. What is your age?

   ____

2. What year did you graduate from medical school?

   ____

3. When did you start practicing medicine at your current clinic?
    - Year ________________________
    - Month ________________________

4. Are you of Hispanic or Latino origin or descent?
    - Yes, Hispanic or Latino
    - No, not Hispanic or Latino

5. What is your race? (Mark one or more)
    - White
    - Black or African American
    - Asian
    - Native Hawaiian or other Pacific Islander
    - American Indian or Alaska Native

- Some other race, ethnicity, or origin
- I would rather not answer

6. Do you currently describe yourself as male, female or transgender?
   - Male
   - Female
   - Transgender
   - None of these
   - I would rather not answer

7. Please describe your relationship status.
   - Single
   - Married
   - In a relationship
   - Living as married
   - Widowed/Widower
   - Divorced or Separated

8. Do you have any children or dependents that you look after?
   - Yes
   - No

   (If yes to above question, then ask)

9. If yes, how many?

____

Quarterly Survey Schedule and Completion Rates

Table A2-4. Quarterly Survey Launch Dates

Quarterly Survey	Launch Date
1	October 3rd, 2019
2	January 8th, 2020
3	April 7th, 2020
4	July 13th, 2020

Table A2-5. Survey Completion Rates

Quarterly Survey	Overall	Condition 1	Condition 2	Condition 3	*p*
1. October, 2019	98.0% (195/199)	98.4% (64/65)	98.4% (63/64)	97.1% (68/70)	1.00
2. January, 2020	91.5% (182/199)	93.8% (61/65)	93.8% (60/64)	87.1% (61/70)	0.32
3. April, 2020	93.0% (185/199)	92.3% (60/65)	90.6% (58/64)	94.3% (66/70)	0.84
4. July, 2020	88.4% (176/199)	90.8% (59/65)	82.8% (53/64)	91.4% (64/70)	0.26

Note: The p-value in the right column is from a Fisher’s exact test, which evaluates whether the completion rates in each of the three conditions are statistically different from one another.

8. Primary Analysis of HM Order Rates

Table A2-6 includes our primary regressions reported in the manuscript using mixed effects binomial logistic regressions. As explained in the statistical analysis section of the manuscript, the model assumes that each patient's number of orders placed follows a binomial distribution, where the number of trials is the patient's number of open topics, and a logit-linear function is used to estimate the probability that a patient has an order placed for any given open topic.

Table A2-6. Estimated Treatment Effects on HM Order Rates (Mixed Effects Binomial Logistic Regressions)

A. Conditions 2 and 3 Combined (vs. Condition 1) Contrast

	Dependent variable:		
	HM Order Rate		
	(1)	(2)	(3)
Condition 2+3 (vs. Condition 1)	0.078 (0.100)	0.078 (0.096)	0.120 (0.082)
Controlling for Patient Baseline HM Completion Rate	No	Yes	Yes
Controlling for Patient and Provider Characteristics	No	No	Yes
Observations	46,631	44,282	44,166

Note: Mixed effects binomial logistic regressions with physician and clinic random effects were used to estimate the treatment effect of Conditions 2 and 3 combined (vs. 1) on HM order rates. Patient characteristics include patients' age, gender, and ZIP code. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Observations differ between models because of variables with missing values. Statistical significance is indicated by: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

B. Condition 2 (vs. Condition 1) and Condition 3 (vs. Condition 1) Contrasts

	Dependent variable:		
	HM Order Rate		
	(1)	(2)	(3)
Condition 2 (vs. Condition 1)	0.078	0.075	0.092
	(0.114)	(0.110)	(0.093)
Condition 3 (vs. Condition 1)	0.078	0.081	0.145
	(0.116)	(0.112)	(0.096)
Controlling for Patient Baseline HM Completion Rate	No	Yes	Yes
Controlling for Patient and Provider Characteristics	No	No	Yes
Observations	46,631	44,282	44,166

Note: Mixed effects binomial logistic regressions with physician and clinic random effects were used to estimate differences between conditions in HM order rates. Patient characteristics include patients' age, gender, and ZIP code. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Observations differ between models because of variables with missing values. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

C. Condition 3 (vs. 2) Contrast

	Dependent variable:		
	HM Order Rate		
	(1)	(2)	(3)
Condition 3 (vs. Condition 2)	0.000	0.006	0.054
	(0.116)	(0.112)	(0.095)

Note: The coefficients reflect linear contrasts using the coefficients from the corresponding models in Panel B above. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

9. Robustness Checks and Secondary Analyses for HM Order Rates

Robustness Check: Pre-registered Alternative Models Estimating Treatment Effects on Order Rates

Table A2-7 shows results from binomial logistic regressions with standard errors clustered by clinic. Again, the model assumes that each patient's number of orders placed follows a binomial distribution, where the number of trials is the patient's number of open topics, and a logit-linear function is used to estimate the probability that a patient has an order placed for any given open topic. Table A2-8 shows linear mixed effects regressions with physician and clinic random effects, and Table A2-9 shows Ordinary Least Squares (OLS) regressions with standard errors clustered by clinic.

Table A2-7. Estimated Treatment Effects on HM Order Rates (Binomial Logistic Regressions with Clustered Standard Errors)

A. Conditions 2 and 3 Combined (vs. Condition 1) Contrast

	Dependent variable:		
	HM Order Rate		
	(1)	(2)	(3)
Condition 2+3 (vs. Condition 1)	0.073	0.069	0.106
	(0.087)	(0.082)	(0.074)
Controlling for Patient Baseline HM Completion Rate	No	Yes	Yes
Controlling for Patient and Provider Characteristics	No	No	Yes
Observations	46,631	44,282	44,166

Note: Binomial logistic regressions with standard errors clustered by clinic were used to estimate the treatment effect of Conditions 2 and 3 combined (vs. 1) on HM order rates. Patient characteristics include patients' age and gender. Zip codes could not be included as a control because the regression did not converge. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Observations differ between models because of variables with missing values. Statistical significance is indicated by: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

B. Condition 2 (vs. Condition 1) and Condition 3 (vs. Condition 1) Contrasts

	Dependent variable:		
	HM Order Rate		
	(1)	(2)	(3)
Condition 2 (vs. Condition 1)	0.102	0.097	0.106
	(0.101)	(0.100)	(0.089)
Condition 3 (vs. Condition 1)	0.043	0.041	0.106
	(0.109)	(0.102)	(0.094)
Controlling for Patient Baseline HM Completion Rate	No	Yes	Yes
Controlling for Patient and Provider Characteristics	No	No	Yes
Observations	46,631	44,282	44,166

Note: Binomial logistic regressions with standard errors clustered by clinic were used to estimate differences between conditions in HM order rates. Patient characteristics include patients' age and gender. Zip code could not be included as a control because the regression did not converge. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Observations differ between models because of variables with missing values. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

C. Condition 3 (vs. 2) Contrast

	Dependent variable:		
	HM Order Rate		
	(1)	(2)	(3)
Condition 3 (vs. Condition 2)	-0.059	-0.056	0.000
	(0.122)	(0.122)	(0.109)

Note: The coefficients reflect linear contrasts using the coefficients from the corresponding models in Panel B above. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

Table A2-8. Estimated Treatment Effects on HM Order Rates (Linear Mixed Effects Regressions)

A. Conditions 2 and 3 Combined (vs. Condition 1) Contrast

	Dependent variable:		
	HM Order Rate		
	(1)	(2)	(3)
Condition 2+3 (vs. Condition 1)	0.008	0.008	0.013*
	(0.009)	(0.009)	(0.007)
Controlling for Patient Baseline HM Completion Rate	No	Yes	Yes
Controlling for Patient and Provider Characteristics	No	No	Yes
Observations	46,631	44,282	44,166

Note: Linear mixed effects regressions with physician and clinic random effects were used to estimate the treatment effect of Conditions 2 and 3 combined (vs. 1) on HM order rates. Patient characteristics include patients' age, gender, and ZIP code. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Observations differ between models because of variables with missing values. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

B. Condition 2 (vs. Condition 1) and Condition 3 (vs. Condition 1) Contrasts

	Dependent variable:		
	HM Order Rate		
	(1)	(2)	(3)
Condition 2 (vs. Condition 1)	0.008	0.008	0.011
	(0.011)	(0.010)	(0.009)
Condition 3 (vs. Condition 1)	0.008	0.009	0.015*
	(0.011)	(0.011)	(0.009)
Controlling for Patient Baseline HM Completion Rate	No	Yes	Yes
Controlling for Patient and Provider Characteristics	No	No	Yes
Observations	46,631	44,282	44,166

Note: Linear mixed effects regressions with physician and clinic random effects were used to estimate differences between conditions in HM order rates. Patient characteristics include patients' age, gender, and ZIP code. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Observations differ between models because of variables with missing values. Statistical significance is indicated by: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

C. Condition 3 (vs. Condition 2) Contrast

	Dependent variable:		
	HM Order Rate		
	(1)	(2)	(3)
Condition 3 (vs. Condition 2)	0.0002	0.001	0.004
	(0.011)	(0.011)	(0.009)

Note: The coefficients reflect linear contrasts using the coefficients from the corresponding models in Panel B above. Statistical significance is indicated by: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

The marginally significant coefficients in Model (3) of Panels A and B are not stable across specifications so we do not interpret them as revealing true treatment effects.

Table A2-9. Estimated Treatment Effects on HM Order Rates (OLS Regressions with Clustered Standard Errors)

A. Conditions 2 and 3 Combined (vs. Condition 1) Contrast

	Dependent variable:		
	HM Order Rate		
	(1)	(2)	(3)
Condition 2+3 (vs. Condition 1)	0.008	0.008	0.014**
	(0.009)	(0.008)	(0.007)
Controlling for Patient Baseline HM Completion Rate	No	Yes	Yes
Controlling for Provider Characteristics	No	No	Yes
Observations	46,631	44,282	44,166
R^2	0.0002	0.004	0.018

Note: OLS regressions with standard errors clustered by clinic were used to estimate the treatment effect of Conditions 2 and 3 combined (vs. 1) on HM order rates. Patient characteristics include patients' age, gender, and ZIP code. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Observations differ between models because of variables with missing values. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

B. Condition 2 (vs. Condition 1) and Condition 3 (vs. Condition 1) Contrasts

	Dependent variable:		
	HM Order Rate		
	(1)	(2)	(3)
Condition 2 (vs. Condition 1)	0.011	0.011	0.015*
	(0.011)	(0.010)	(0.008)
Condition 3 (vs. Condition 1)	0.005	0.005	0.013
	(0.011)	(0.010)	(0.009)
Controlling for Patient Baseline HM Completion Rate	No	Yes	Yes
Controlling for Patient and Provider Characteristics	No	No	Yes
Observations	46,631	44,282	44,166
R^2	0.0003	0.004	0.018

Note: OLS regressions with standard errors clustered by clinic were used to estimate differences between conditions in HM order rates. Patient characteristics include patients' age, gender, and ZIP code. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Observations differ between models because of variables with missing values. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

C. Condition 3 (vs. Condition 2) Contrast

	Dependent variable:		
	HM Order Rate		
	(1)	(2)	(3)
Condition 3 (vs. Condition 2)	-0.006	-0.006	-0.002
	(0.012)	(0.012)	(0.008)

Note: The coefficients reflect linear contrasts using the coefficients from the corresponding models in Panel B above. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

The significant and marginally significant coefficients in Model (3) of Panels A and B are not stable across specifications so we do not interpret them as revealing true treatment effects.

Secondary Analysis: Order Rates Moderated by Physician Baseline Performance

We examined whether the effects of peer comparison on order rates were moderated by baseline performance (physicians' HM completion rates at baseline; from July-October 2019) using both a continuous and categorical version of the moderator (Tables A2-10 and A2-11). We were specifically interested in the Condition 2 (vs. Condition 1) contrast and its interaction with baseline performance because they allow us to isolate the heterogeneous treatment effects of peer comparison information (without conflation with the potential heterogeneous effects of leadership support training).

Table A2-10. Order Rates Moderated by Physicians' Baseline Performance

(Continuous)

	Dependent variable:	
	HM Order Rate	
	(1)	(2)
Condition 2 (vs. Condition 1)	0.097	0.083
	(0.100)	(0.092)
Condition 3 (vs. Condition 1)	0.058	0.091
	(0.102)	(0.095)
Baseline HM Completion Rate	0.022***	0.013**
	(0.006)	(0.005)
Condition 2 (vs. Condition 1) * Baseline HM Completion Rate	-0.005 (0.007)	-0.002 (0.006)
Condition 3 (vs. Condition 1) * Baseline HM Completion Rate	0.000 (0.007)	0.004 (0.006)
Controlling for Patient and Provider Characteristics	No	Yes
Observations	46,336	46,218

Note: Mixed effects binomial logistic regressions with physician and clinic random effects were used to estimate the coefficients. Baseline HM Completion Rate is mean-centered. Patient characteristics include patients' age, gender, and ZIP code. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Observations

---

differ between models because of variables with missing values. Statistical significance is indicated by: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

Table A2-11. Order Rates Moderated by Baseline Performance (Categorical)

A. Regressions with Interactions

	Dependent variable:	
	HM Order Rate	
	(1)	(2)
Condition 2 (vs. Condition 1)	0.133	0.004
	(0.167)	(0.151)
Condition 3 (vs. Condition 1)	0.068	-0.020
	(0.169)	(0.151)
Baseline “Almost High Performer”	0.209	-0.001
	(0.144)	(0.126)
Baseline “High Performer”	0.356**	0.167
	(0.149)	(0.133)
Baseline “Top Performer”	0.625***	0.311*
	(0.196)	(0.178)
Condition 2 (vs. Condition 1) *	-0.036	0.144
Baseline “Almost High Performer”	(0.194)	(0.173)
Condition 2 (vs. Condition 1) *	-0.064	0.045
Baseline “High Performer”	(0.198)	(0.174)
Condition 2 (vs. Condition 1) *	-0.066	0.043
Baseline “Top Performer”	(0.260)	(0.227)
Condition 3 (vs. Condition 1) *	-0.109	0.112
Baseline “Almost High Performer”	(0.190)	(0.167)
Condition 3 (vs. Condition 1) *	0.028	0.158
Baseline “High Performer”	(0.202)	(0.175)
Condition 3 (vs. Condition 1) *	0.107	0.220
Baseline “Top Performer”	(0.252)	(0.222)
Controlling for Patient and Provider Characteristics	No	Yes
Observations	46,336	46,218

Note: Mixed effects binomial logistic regressions with physician and clinic random effects were used to estimate the coefficients. The baseline performance tiers include: “Almost High Performer” = 55-65% completion rate; “High Performer” = >65% completion rate, but not top 25 ranked score; “Top Performer” = top 25 ranked score; with “Low Performer = <55% completion rate as the reference group. Patient characteristics include patients’ age, gender, and ZIP code. Provider characteristics include providers’ gender, race, years since graduating medical school, and years of working at UCLA Health. Observations differ between models because of variables with missing values. Statistical significance is indicated by: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

B. Treatment Effects within Each Performance Tier

	Dependent variable:			
	HM Order Rates			
	(1)	(2)	(3)	(4)
Condition 2 (vs. Condition 1)	0.004	0.149	0.050	0.047
	(0.151)	(0.125)	(0.132)	(0.196)
Condition 3 (vs. Condition 1)	-0.020	0.092	0.138	0.200
	(0.151)	(0.129)	(0.136)	(0.187)
Baseline Performance Tier	Low Performer	Almost High Performer	High Performer	Top Performer

Note: The coefficients reflect the estimated treatment effects among PCPs within each performance tier (indicated in the last row of the table) and come from Model (2) in Panel A. For instance, the treatment effect of Condition 2 (vs. 1) within the "Top Performer" tier (0.047) is estimated using the following linear contrast with the coefficients from Model (2) in Panel A: Condition 2 (vs. Condition 1) * Baseline "Top Performer" + Condition 2 (vs. Condition 1) = (0.043 + 0.004). The baseline performance tiers include: "Low Performer" = <55% completion rate; "Almost High Performer" = 55-65% completion rate; "High Performer" = >65% completion rate, but not top 25 ranked score; "Top Performer" = top 25 ranked score. Statistical significance is indicated by: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

10. Primary Analysis of Job Satisfaction and Burnout in April 2020

Table A2-12. Estimated Treatment Effects on Job Satisfaction (April 2020)

A. Condition 2 (vs. Condition 1) and Condition 3 (vs. Condition 1) Contrasts

	Dependent variable:			
	Job Satisfaction			
	(1)	(2)	(3)	(4)
Condition 2 (vs. Condition 1)	-0.519**	-0.631***	-0.549**	-0.564**
	(0.250)	(0.217)	(0.235)	(0.240)
Condition 3 (vs. Condition 1)	-0.179	-0.237	-0.103	-0.120
	(0.207)	(0.148)	(0.155)	(0.150)
Controlling for Baseline Job Satisfaction	No	Yes	Yes	Yes
Controlling for Provider Characteristics	No	No	Yes	Yes
Controlling for COVID Case Load	No	No	No	Yes
Observations	183	177	177	177
R^2	0.028	0.361	0.398	0.415

Note: OLS regressions with standard errors clustered by clinic were used to estimate differences between conditions in job satisfaction. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Observations differ between models because of variables with missing values. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

B. Condition 3 (vs. Condition 2) Contrast

	Dependent variable:			
	Job Satisfaction			
	(1)	(2)	(3)	(4)
Condition 3 (vs. Condition 2)	0.341	0.395**	0.447**	0.444*

	(0.249)	(0.197)	(0.219)	(0.230)

Note: The coefficients reflect linear contrasts using the coefficients from the corresponding models in Panel A above. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

Table A2-13. Estimated Treatment Effects on Burnout (April 2020)

A. Condition 2 (vs. Condition 1) and Condition 3 (vs. Condition 1) Contrasts

	Dependent variable:			
	Burnout			
	(1)	(2)	(3)	(4)
Condition 2 (vs. Condition 1)	0.539**	0.360**	0.329**	0.330**
	(0.217)	(0.152)	(0.151)	(0.152)
Condition 3 (vs. Condition 1)	0.159	-0.070	-0.112	-0.111
	(0.141)	(0.112)	(0.123)	(0.121)
Controlling for Baseline Burnout	No	Yes	Yes	Yes
Controlling for Provider Characteristics	No	No	Yes	Yes
Controlling for COVID Case Load	No	No	No	Yes
Observations	180	174	174	174
R^2	0.065	0.471	0.486	0.488

Note: OLS regressions with standard errors clustered by clinic were used to estimate differences between conditions in burnout. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Observations differ between models because of variables with missing values. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

B. Condition 3 (vs. Condition 2) Contrast

	Dependent variable:			
	Burnout			
	(1)	(2)	(3)	(4)
Condition 3 (vs. Condition 2)	-0.380*	-0.430**	-0.441**	-0.441**
	(0.211)	(0.172)	(0.181)	(0.183)

Note: The coefficients reflect linear contrasts using the coefficients from the corresponding models in Panel A above. Statistical significance is indicated by: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

11. Robustness Checks and Secondary Analyses for Job Satisfaction and Burnout

Placebo Test: Regression Analysis of Proficiency with CareConnect

We estimated our primary specification with an outcome that we would not expect to be impacted by the interventions. We specifically used physicians' responses to the following item in the April 2020 quarterly survey as the outcome: "My proficiency with using CareConnect is: (1) Poor, (2) Marginal, (3) Satisfactory, (4) Good, (5) Optimal."

Table A2-14. Estimated Treatment Effects on Proficiency with CareConnect (April 2020)

A. Condition 2 (vs. Condition 1) and Condition 3 (vs. Condition 1) Contrasts

	Dependent variable:
	CareConnect Proficiency
	(1)
Condition 2 (vs. Condition 1)	-0.139
	(0.087)
Condition 3 (vs. Condition 1)	-0.168**
	(0.082)
Controlling for Baseline CareConnect Proficiency	Yes
Controlling for Provider Characteristics	Yes
Observations	179
R^2	0.320

Note: OLS regression with standard errors clustered by clinic was used to estimate differences between conditions in CareConnect proficiency. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Statistical significance is indicated by: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

B. Condition 3 (vs. Condition 2) Contrast

	Dependent variable:
	CareConnect Proficiency
	(3)
Condition 3 (vs. Condition 2)	-0.029
	(0.093)

Note: The coefficient reflects a linear contrast using the coefficients from Panel A above. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

The placebo item was not impacted by either intervention: the effect of the peer comparison intervention alone (Condition 2 (vs. 1)) was null, and the effect of leadership training was null (Condition 3 (vs. 2)). We do not know why the two interventions combined (Condition 3 (vs. 1)) had a negative effect on the placebo item, and we suspect this is spurious. Importantly, this effect could not explain our findings about job satisfaction and burnout.

Secondary Analysis: Effects on Physician Leads and Non-Leads

To better understand the effects of leadership support training, we also examined whether the benefits trickled down to fellow PCPs who were not physician leads and thus did not receive training personally. Specifically, we analyzed the effects of our interventions on PCPs who were not leaders [“non-leads”; columns (1)-(3)] and PCPs who were leaders [“leads”; columns (4)-(6)]. Note that we use Condition 2 as the reference group in these regressions because the main contrast of interest is Condition 3 (vs. Condition 2), which reflects the impact of leadership training.

Table A2-15. Estimated Treatment Effects on Job Satisfaction for Physician Leads and Non-Leads (April 2020)

	Dependent variable:					
	Job Satisfaction					
	(1)	(2)	(3)	(4)	(5)	(6)
Condition 1 (vs. Condition 2)	0.542*	0.536*	0.524	0.505	0.745	0.463
	(0.299)	(0.296)	(0.330)	(0.569)	(0.569)	(0.614)
Condition 3 (vs. Condition 2)	0.294	0.356	0.435	0.636	0.636	0.946*
	(0.329)	(0.348)	(0.362)	(0.498)	(0.437)	(0.543)
Subsample	Only Non-Leads	Only Non-Leads	Only Non-Leads	Only Leads	Only Leads	Only Leads
Controlling for Baseline Job Satisfaction	No	Yes	Yes	No	Yes	Yes
Controlling for Provider Characteristics	No	No	Yes	No	No	Yes
Observations	152	147	147	31	31	31
R^2	0.030	0.056	0.135	0.060	0.161	0.488

Note: OLS regressions with standard errors clustered by clinic were used to estimate the differences between conditions in job satisfaction. Provider characteristics include providers’ gender, race, years since graduating medical school, and years of working at UCLA Health. Statistical significance is indicated by: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

Table A2-16. Estimated Treatment Effects on Burnout for Physician Leads and Non-Leads (April 2020)

	Dependent variable:					
	Burnout					
	(1)	(2)	(3)	(4)	(5)	(6)
Condition 1 (vs. Condition 2)	-0.511**	-0.520**	-0.518**	-0.744**	-0.775**	-0.852**
	(0.249)	(0.233)	(0.254)	(0.322)	(0.302)	(0.363)
Condition 3 (vs. Condition 2)	-0.419	-0.476*	-0.469*	-0.209	-0.210	-0.453
	(0.256)	(0.264)	(0.282)	(0.301)	(0.305)	(0.446)
Subsample	Only Non-Leads	Only Non-Leads	Only Non-Leads	Only Leads	Only Leads	Only Leads
Controlling for Baseline Burnout	No	Yes	Yes	No	Yes	Yes
Controlling for Provider Characteristics	No	No	Yes	No	No	Yes
Observations	150	145	145	30	30	30
R^2	0.059	0.082	0.102	0.173	0.177	0.360

Note: OLS regressions with standard errors clustered by clinic were used to estimate the differences between conditions in burnout. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Statistical significance is indicated by: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

Secondary Analysis: Job Satisfaction and Burnout Moderated by Baseline Performance

We examined whether the effects of peer comparison on job satisfaction and burnout were moderated by baseline performance (physicians' HM completion rates at baseline; from July-October 2019) using both a continuous and categorical version of the moderator (Tables A2-17 and A2-18). We were specifically interested in the Condition 2 (vs. Condition 1) contrast and its interaction with baseline performance because they allow us to isolate the heterogeneous treatment effects of peer comparison information (without conflation with the potential heterogeneous effects of leadership support training).

Table A2-17. Job Satisfaction Moderated by Baseline Performance (Continuous)

	Dependent variable:		
	Job Satisfaction		
	(1)	(2)	(3)
Condition 2 (vs. Condition 1)	-0.507^{**}	-0.605^{***}	-0.524^{**}
	(0.243)	(0.218)	(0.230)
Condition 3 (vs. Condition 1)	-0.170	-0.215	-0.090
	(0.196)	(0.143)	(0.153)
Baseline HM Completion Rate	-0.020^{**}	-0.005	-0.006
	(0.010)	(0.008)	(0.007)
Condition 2 (vs. Condition 1) * Baseline HM Completion Rate	0.031^{*}	0.019	0.017
	(0.016)	(0.015)	(0.013)
Condition 3 (vs. Condition 1) * Baseline HM Completion Rate	0.034^{***}	0.012	0.014
	(0.013)	(0.011)	(0.010)
Controlling for Baseline Job Satisfaction	No	Yes	Yes
Controlling for Provider Characteristics	No	No	Yes
Observations	182	176	176
R^2	0.044	0.371	0.409

Note: OLS regressions with standard errors clustered by clinic were used to estimate the coefficients. Baseline HM Completion Rate is mean-centered. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Observations differ between models because of variables with missing values. Statistical significance is indicated by: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$.

Table A2-18. Burnout Moderated by Baseline Performance (Continuous)

	Dependent variable:		
	Burnout		
	(1)	(2)	(3)
Condition 2 (vs. Condition 1)	0.533**	0.349**	0.315**
	(0.213)	(0.152)	(0.150)
Condition 3 (vs. Condition 1)	0.153	-0.083	-0.135
	(0.131)	(0.110)	(0.128)
Baseline HM Completion Rate	0.014	0.005	0.009
	(0.011)	(0.006)	(0.007)
Condition 2 (vs. Condition 1) * Baseline HM Completion Rate	-0.007	-0.007	-0.008
	(0.016)	(0.009)	(0.010)
Condition 3 (vs. Condition 1) * Baseline HM Completion Rate	-0.006	-0.003	-0.005
	(0.012)	(0.008)	(0.008)
Controlling for Baseline Burnout	No	Yes	Yes
Controlling for Provider Characteristics	No	No	Yes
Observations	179	173	173
R^2	0.078	0.473	0.492

Note: OLS regressions with standard errors clustered by clinic were used to estimate the coefficients. Baseline HM Completion Rate is mean-centered. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Observations differ between models because of variables with missing values. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

Table A2-19. Job Satisfaction Moderated by Baseline Performance Tier (Categorical)

A. Regressions with Interactions

	Dependent variable:		
	Job Satisfaction		
	(1)	(2)	(3)
Condition 2 (vs. Condition 1)	-1.250^{***}	-1.020^{***}	-0.886^{***}
	(0.247)	(0.193)	(0.198)
Condition 3 (vs. Condition 1)	-0.900^{***}	-0.440^{**}	-0.352^{*}
	(0.252)	(0.188)	(0.208)
Baseline “Almost High Performer”	-0.700^{**}	-0.336	-0.333
	(0.269)	(0.217)	(0.204)
Baseline “High Performer”	-0.762^{***}	-0.274	-0.313
	(0.277)	(0.236)	(0.293)
Baseline “Top Performer”	-0.600^{***}	-0.080	-0.227
	(0.229)	(0.735)	(0.584)
Condition 2 (vs. Condition 1) * Baseline “Almost High Performer”	0.950	0.585	0.448
	(0.613)	(0.440)	(0.479)
Condition 2 (vs. Condition 1) * Baseline “High Performer”	1.130^{**}	0.569	0.551
	(0.523)	(0.461)	(0.499)
Condition 2 (vs. Condition 1) * Baseline “Top Performer”	0.707^{*}	0.376	0.371
	(0.403)	(0.799)	(0.696)
Condition 3 (vs. Condition 1) * Baseline “Almost High Performer”	0.700^{*}	0.240	0.272
	(0.366)	(0.265)	(0.278)
Condition 3 (vs. Condition 1) * Baseline “High Performer”	1.329^{***}	0.450	0.446
	(0.505)	(0.355)	(0.388)
Condition 3 (vs. Condition 1) * Baseline “Top Performer”	0.955^{**}	0.103	0.345
	(0.428)	(0.792)	(0.643)
Controlling for Baseline Job Satisfaction	No	Yes	Yes
Controlling for Provider Characteristics	No	No	Yes
Observations	182	176	176

R^2	0.063	0.374	0.411

Note: OLS regressions with standard errors clustered by clinic were used to estimate the coefficients. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Observations differ between models because of variables with missing values. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

B. Treatment Effects within Each Performance Tier

	Dependent variable:			
	Job Satisfaction			
	(1)	(2)	(3)	(4)
Condition 2 (vs. Condition 1)	-0.886^{***}	-0.438	-0.335	-0.515
	(0.198)	(0.414)	(0.459)	(0.719)
Condition 3 (vs. Condition 1)	-0.352^{*}	-0.081	0.093	-0.008
	(0.208)	(0.222)	(0.343)	(0.678)
Baseline Performance Tier	Low Performer	Almost High Performer	High Performer	Top Performer

Note: The coefficients reflect the estimated treatment effects among PCPs within each performance tier (indicated in the last row of the table) and come from Model (3) in Panel A. For instance, the treatment effect of Condition 2 (vs. 1) within the "Top Performer" tier (-0.515) is estimated using the following linear contrast with the coefficients from Model (3) in Panel A: Condition 2 (vs. Condition 1) * Baseline "Top Performer" + Condition 2 (vs. Condition 1) = (0.371 - 0.886). The baseline performance tiers include: "Low Performer" = <55% completion rate; "Almost High Performer" = 55-65% completion rate; "High Performer" = >65% completion rate, but not top 25 ranked score; "Top Performer" = top 25 ranked score. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

Table A2-20. Burnout Moderated by Baseline Performance (Categorical)

A. Regressions with Interactions

	Dependent variable:		
	Burnout		
	(1)	(2)	(3)
Condition 2 (vs. Condition 1)	0.731**	0.533***	0.485***
	(0.309)	(0.164)	(0.168)
Condition 3 (vs. Condition 1)	0.231	-0.125	-0.164
	(0.206)	(0.177)	(0.189)
Baseline “Almost High Performer”	0.181	0.026	0.112
	(0.202)	(0.108)	(0.116)
Baseline “High Performer”	0.183	0.002	0.085
	(0.261)	(0.125)	(0.129)
Baseline “Top Performer”	0.631	0.428**	0.669**
	(0.538)	(0.203)	(0.283)
Condition 2 (vs. Condition 1) * Baseline “Almost High Performer”	-0.347	-0.333	-0.289
	(0.426)	(0.249)	(0.265)
Condition 2 (vs. Condition 1) * Baseline “High Performer”	-0.271	-0.274	-0.248
	(0.420)	(0.238)	(0.235)
Condition 2 (vs. Condition 1) * Baseline “Top Performer”	-0.274	-0.195	-0.294
	(0.597)	(0.312)	(0.389)
Condition 3 (vs. Condition 1) * Baseline “Almost High Performer”	-0.231	-0.059	-0.090
	(0.244)	(0.163)	(0.160)
Condition 3 (vs. Condition 1) * Baseline “High Performer”	0.217	0.415**	0.431**
	(0.331)	(0.178)	(0.174)
Condition 3 (vs. Condition 1) * Baseline “Top Performer”	-0.531	-0.495*	-0.640**
	(0.578)	(0.261)	(0.282)
Controlling for Baseline Burnout	No	Yes	Yes

Controlling for Provider Characteristics	No	No	Yes
Observations	179	173	173
R^2	0.098	0.511	0.532

Note: OLS regressions with standard errors clustered by clinic were used to estimate the coefficients. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Observations differ between models because of variables with missing values. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

B. Treatment Effects within Each Performance Tier

	Dependent variable:			
	Burnout			
	(1)	(2)	(3)	(4)
Condition 2 (vs. Condition 1)	0.485^{***}	0.196	0.237	0.191
	(0.168)	(0.254)	(0.226)	(0.387)
Condition 3 (vs. Condition 1)	-0.164	-0.254	0.267^{*}	-0.805^{***}
	(0.189)	(0.156)	(0.157)	(0.285)
Baseline Performance Tier	Low Performer	Almost High Performer	High Performer	Top Performer

Note: The coefficients reflect the estimated treatment effects among PCPs within each performance tier (indicated in the last row of the table) and come from Model (3) in Panel A. For instance, the treatment effect of Condition 2 (vs. 1) within the "Top Performer" tier (0.191) is estimated using the following linear contrast with the coefficients from Model (3) in Panel A: Condition 2 (vs. Condition 1) * Baseline "Top Performer" + Condition 2 (vs. Condition 1) = (0.485 - 0.294). The baseline performance tiers include: "Low Performer" = <55% completion rate; "Almost High Performer" = 55-65% completion rate; "High Performer" = >65% completion rate, but not top 25 ranked score; "Top Performer" = top 25 ranked score. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

12. Treatment Effect Persistence

Table A2-21. Estimated Long-term Treatment Effects on Job Satisfaction (July 2020)

A. Condition 2 (vs. Condition 1) and Condition 3 (vs. Condition 1) Contrasts

	Dependent variable:		
	Job Satisfaction		
	(1)	(2)	(3)
Condition 2 (vs. Condition 1)	-0.579**	-0.748***	-0.601**
	(0.286)	(0.242)	(0.247)
Condition 3 (vs. Condition 1)	-0.014	-0.169	0.014
	(0.251)	(0.209)	(0.223)
Controlling for Baseline Job Satisfaction	No	Yes	Yes
Controlling for Provider Characteristics	No	No	Yes
Observations	175	170	170
R^2	0.038	0.358	0.412

Note: OLS regressions with standard errors clustered by clinic were used to estimate differences between conditions in job satisfaction. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Observations differ between models because of variables with missing values. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

B. Condition 3 (vs. Condition 2) Contrast

	Dependent variable:		
	Job Satisfaction		
	(1)	(2)	(3)
Condition 3 (vs. Condition 2)	0.565*	0.579**	0.615**
	(0.327)	(0.247)	(0.245)

Note: The coefficients reflect linear contrasts using the coefficients from the corresponding models in Panel A above. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

Table A2-22. Estimated Long-term Treatment Effects on Burnout (July 2020)

A. Condition 2 (vs. Condition 1) and Condition 3 (vs. Condition 1) Contrasts

	Dependent variable:		
	Burnout		
	(1)	(2)	(3)
Condition 2 (vs. Condition 1)	0.380*	0.212*	0.088
	(0.207)	(0.124)	(0.148)
Condition 3 (vs. Condition 1)	0.176	-0.041	-0.123
	(0.193)	(0.167)	(0.192)
Controlling for Baseline Burnout	No	Yes	Yes
Controlling for Provider Characteristics	No	No	Yes
Observations	172	168	168
R^2	0.027	0.418	0.492

Note: OLS regressions with standard errors clustered by clinic were used to estimate differences between conditions in burnout. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Observations differ between models because of variables with missing values. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

B. Condition 3 (vs. Condition 2) Contrast

	Dependent variable:		
	Burnout		
	(1)	(2)	(3)
Condition 3 (vs. Condition 2)	-0.203	-0.252	-0.211
	(0.239)	(0.188)	(0.215)

Note: The coefficients reflect linear contrasts using the coefficients from the corresponding models in Panel A above. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

13. Regression Analysis of Perceived Leadership Support

Table A2-23. Estimated Treatment Effects on Perceived Leadership Support (April 2020)

A. Condition 2 (vs. Condition 1) and Condition 3 (vs. Condition 1) Contrasts

	Dependent variable:			
	Perceived Leadership Support			
	(1)	(2)	(3)	(4)
Condition 2 (vs. Condition 1)	-0.499*	-0.635***	-0.599**	-0.599**
	(0.275)	(0.222)	(0.237)	(0.240)
Condition 3 (vs. Condition 1)	0.029	-0.099	-0.040	-0.039
	(0.251)	(0.156)	(0.162)	(0.164)
Controlling for Baseline Perceived Leadership Support	No	Yes	Yes	Yes
Controlling for Provider Characteristics	No	No	Yes	Yes
Controlling for COVID Case Load	No	No	No	Yes
Observations	184	179	179	179
R^2	0.049	0.398	0.434	0.434

Note: OLS regressions with standard errors clustered by clinic were used to estimate differences between conditions in perceived leadership support. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Observations differ between models because of variables with missing values. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

B. Condition 3 (vs. Condition 2) Contrast

	Dependent variable:			
	Perceived Leadership Support			
	(1)	(2)	(3)	(4)
Condition 3 (vs. Condition 2)	0.528*	0.536**	0.560**	0.560**
	(0.282)	(0.233)	(0.238)	(0.242)

Note: The coefficients reflect linear contrasts using the coefficients from the corresponding models in Panel A above. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

Table A2-24. Estimated Long-term Treatment Effects on Perceived Leadership Support (July 2020)

A. Condition 2 (vs. Condition 1) and Condition 3 (vs. Condition 1) Contrasts

	Dependent variable:		
	Perceived Leadership Support		
	(1)	(2)	(3)
Condition 2 (vs. Condition 1)	-0.590**	-0.740***	-0.689***
	(0.276)	(0.222)	(0.219)
Condition 3 (vs. Condition 1)	-0.077	-0.275*	-0.199
	(0.246)	(0.166)	(0.174)
Controlling for Baseline Perceived Leadership Support	No	Yes	Yes
Controlling for Provider Characteristics	No	No	Yes
Observations	175	171	171
R^2	0.055	0.437	0.451

Note: OLS regressions with standard errors clustered by clinic were used to estimate differences between conditions in perceived leadership support. Provider characteristics include providers' gender, race, years since graduating medical school, and years of working at UCLA Health. Observations differ between models because of variables with missing values. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

B. Condition 3 (vs. Condition 2) Contrast

	Dependent variable:		
	Perceived Leadership Support		
	(1)	(2)	(3)
Condition 3 (vs. Condition 2)	0.513	0.465*	0.489*
	(0.316)	(0.251)	(0.251)

Note: The coefficients reflect linear contrasts using the coefficients from the corresponding models in Panel A above. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

We also examined whether the effects of leadership training on perceived leadership support trickled down to fellow PCPs who were not physician leads and thus did not receive training personally. Specifically, we analyzed the effects of our interventions on PCPs who were not leaders [“non-leads”; columns (1)-(3)] and PCPs who were leaders [“leads”; columns (4)-(6)]. Again, we use Condition 2 as the reference group in these regressions because the main contrast of interest is Condition 3 (vs. Condition 2), which reflects the impact of leadership training.

Table A2-25. Estimated Treatment Effects on Perceived Leadership Support for Physician Leads and Non-Leads (April 2020)

	Dependent variable:					
	Perceived Leadership Support					
	(1)	(2)	(3)	(4)	(5)	(6)
Condition 1 (vs. Condition 2)	0.431	0.430	0.491	0.909^{*}	0.967^{*}	0.565
	(0.279)	(0.285)	(0.323)	(0.477)	(0.503)	(0.547)
Condition 3 (vs. Condition 2)	0.491^{*}	0.541^{*}	0.607^{*}	0.727	0.727	0.990
	(0.279)	(0.296)	(0.315)	(0.529)	(0.533)	(0.586)
Subsample	Only Non-Leads	Only Non-Leads	Only Non-Leads	Only Leads	Only Leads	Only Leads
Controlling for Baseline Perceived Leadership Support	No	Yes	Yes	No	Yes	Yes
Controlling for Provider Characteristics	No	No	Yes	No	No	Yes
Observations	153	148	148	31	31	31
R^2	0.041	0.062	0.163	0.122	0.128	0.418

Note: OLS regressions with standard errors clustered by clinic were used to estimate differences between conditions in perceived leadership support. Provider characteristics include providers’ gender, race, years since graduating medical school, and years of working at UCLA Health. Statistical significance is indicated by: *** p < 0.01; ** p < 0.05; * p < 0.10.

14. Coding of Open-Ended Responses

The follow-up survey (conducted in April 2021) showed all the PCPs, regardless of their original experimental condition, an example of the peer comparison email. Then PCPs were asked, “Would you prefer that the Department resumes sending these types of emails to physicians?” The responses were qualitatively coded by two PCPs, who had the necessary contextual knowledge on how clinical care is practiced at the DOM primary care network. They were blind to the hypotheses, design of the experiment, and survey respondents’ study conditions. The coders categorized the responses based on whether the PCPs expressed any negative reaction and, more specifically, whether they indicated that the peer comparison information would be harmful (e.g., offensive, stress-inducing). After confirming that the responses had high interrater reliability (negative reaction, Cohen’s Kappa = 0.83; harm, Cohen’s Kappa = 0.77), the coders reconciled the remaining differences in their categorizations through discussion. According to their final ratings, 35.3% of PCPs reacted negatively to the peer comparison information, and 14.1% of PCPs went as far as to indicate that it would be harmful.

15. Perceived Control over HM Completion Rates

The follow-up survey (conducted in April 2021), asked for agreement (1- Strongly Disagree; 5- Strongly Agree) with the following item:

> *Physicians can improve their Health Maintenance completion rate with enough effort.*

Table A2-26. Distribution and Summary Statistics of Perceived Control Item

A. Distribution

Response	Count (Frequency)
Strongly disagree (1)	7 (4.6%)
Disagree	27 (17.8%)
Neither agree nor disagree	44 (29.0%)
Agree	64 (42.1%)
Strongly agree (5)	10 (6.6%)

B. Summary Statistics by Baseline Performance Tier

Statistic	All Respondents	Low Performers	Almost High Performers	High Performers	Top Performers
Mean (SD)	3.3 (1.0)	3.1 (1.1)	3.3 (1.0)	3.3 (1.0)	3.7 (0.7)
Median (Q1, Q3)	3 (3, 4)	3 (2, 4)	3 (2, 4)	4 (3, 4)	4 (3, 4)

Note that responses to the perceived control item were predicted by PCPs' baseline performance tier (*p* = 0.010; estimated from a linear regression with tier treated as a continuous variable, with values ranging from 1 = Low Performers to 4 = Top Performers, and with clustered SEs at the clinic level).

REFERENCES

Allcott, Hunt and Judd B. Kessler (2019), "The welfare effects of nudges: A case study of energy use social comparisons," *American Economic Journal: Applied Economics*, 11 (1), 236–76.

Allcott, Hunt and Todd Rogers (2014), "The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation," *American Economic Review*, 104 (10), 3003–37.

Ashraf, Nava, Oriana Bandiera, and Scott S. Lee (2014), "Awards unbundled: Evidence from a natural field experiment," *Journal of Economic Behavior and Organization*, 100, 44–63.

Azmat, Ghazala and Nagore Iriberri (2010), "The importance of relative performance feedback information: Evidence from a natural experiment using high school students," *Journal of Public Economics*, 94 (7–8), 435–52.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2013), "Team incentives: Evidence from a firm level experiment," *Journal of the European Economic Association*, 11 (5), 1079–1114.

Barankay, Iwan (2012), "Rank incentives: Evidence from a randomized workplace experiment," *Business Economics and Public Policy Papers University, Working Paper, University of Pennsylvania*, 1–40.

Blader, Steven, Claudine Gartenberg, and Andrea Prat (2020), "The contingent effect of management practices," *The Review of Economic Studies*, 87 (2), 721–49.

Bobbio, Andrea, Maria Bellan, and Anna Maria Manganelli (2012), "Empowering leadership, perceived organizational support, trust, and job burnout for nurses: A study in an Italian general hospital," *Health Care Management Review*, 37 (1), 77–87.

Bogard, Jonathan E., Magali A. Delmas, Noah J. Goldstein, and I. Stephanie Vezich (2020), "Target, distance, and valence: Unpacking the effects of normative feedback," *Organizational Behavior and Human Decision Processes*, 161 (S), 61–73.

Brown, Douglas J., D. Lance Ferris, Daniel Heller, and Lisa M. Keeping (2007), "Antecedents and consequences of the frequency of upward and downward social comparisons at work," *Organizational Behavior and Human Decision Processes*, 102 (1), 59–75.

Buntinx, F., J. A. Knottnerus, H. F.J.M. Crebolder, T. Seegers, G. G.M. Essed, and H. Schouten (1993), "Does feedback improve the quality of cervical smears? A randomized controlled trial," *British Journal of General Practice*, 43 (370), 194–98.

Bursztyn, Leonardo and Robert Jensen (2015), "How does peer pressure affect educational investments?," *The Quarterly Journal of Economics*, 130 (3), 1329–67.

Butera, Luigi, Robert Metcalfe, William Morrison, and Dmitry Taubinsky (2022), "Measuring the welfare effects of shame and pride," *American Economic Review*, 112 (1), 122–68.

"Care Compare: Doctors and Clinicians Initiative" (2021), *CMS.gov Centers for Medicare & Medicaid Services*, [available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Care-Compare-DAC-Initiative].

Clee, Mona A. and Robert A. Wicklund (1980), "Consumer Behavior and Psychological Reactance," *Journal of Consumer Research*, 6 (4), 389.

Dmitrienko, Alex and Ajit C. Tamhane (2007), "Gatekeeping procedures with clinical trial applications," *Pharmaceutical Statistics*, 6 (3), 171–80.

Doctor, Jason N., Andy Nguyen, Roneet Lev, Jonathan Lucas, Tara Knight, Henu Zhao, and Michael Menchine (2018), "Opioid prescribing decreases after learning of a patient's fatal overdose," *Science*, 361 (6402), 588–90.

Dolan, Emily D., David Mohr, Michele Lempa, Sandra Joos, Stephan D. Fihn, Karin M. Nelson, and Christian D. Helfrich (2015), “Using a single item to measure burnout in primary care staff: A psychometric evaluation,” *Journal of General Internal Medicine*, 30 (5), 582–87.

Dolbier, Christyn L., Judith A. Webster, Katherine T. McCalister, Mark W. Mallon, and Mary A. Steinhardt (2005), “Reliability and validity of a single-item measure of job satisfaction,” *American Journal of Health Promotion*, 19 (3), 194–98.

Fox, Craig R., Jason N. Doctor, Noah J. Goldstein, Daniella Meeker, Stephen D. Persell, and Jeffrey A. Linder (2020), “Details matter: Predicting when nudging clinicians will succeed or fail,” *BMJ*, 370, 1–3.

Frey, Bruno S. and Stephan Meier (2004), “Social comparisons and pro-social behavior: Testing conditional cooperation in a field experiment,” *American Economic Review*, 94 (5), 1717–22.

Gallus, Jana, Joseph Reiff, Emir Kamenica, and Alan Page Fiske (2021), “Relational incentives theory,” *Psychological Review*.

George, Jennifer M. and Gareth R. Jones (1996), “The experience of work and turnover intentions: Interactive effects of value attainment, job satisfaction, and positive mood,” *Journal of Applied Psychology*, 81 (3), 318–25.

Gerber, Alan S. and Todd Rogers (2009), “Descriptive social norms and motivation to vote: Everybody’s voting and so should you,” *Journal of Politics*, 71 (1), 178–91.

Han, Shasha, Tait D. Shanafelt, Christine A. Sinsky, Karim M. Awad, Liselotte N. Dyrbye, Lynne C. Fiscus, Mickey Trockel, and Joel Goh (2019), “Estimating the attributable cost of physician burnout in the United States,” *Annals of Internal Medicine*, 170 (11), 784–90.

Hassol, Andrea, Nathan West, Jessie Gerteis, and Morgan Michaels (2021), “Evaluation of the oncology care model.”

Hennig-Schmidt, Heike, Abdolkarim Sadrieh, and Bettina Rockenbach (2010), “In search of workers’ real effort reciprocity-a field and a laboratory experiment,” *Journal of the European Economic Association*, 8 (4), 817–37.

Hermes, Henning, Martin Huschens, Franz Rothlauf, and Daniel Schunk (2021), “Motivating low-achievers—Relative performance feedback in primary schools,” *Journal of Economic Behavior and Organization*, 187, 45–59.

Judge, Timothy A, Carl J Thoresen, Joyce E Bono, and Gregory K Patton (2001), “The job satisfaction–job performance relationship: A qualitative and quantitative review.,” *Psychological Bulletin*, 127 (3), 376–407.

Kane, Leslie (2021), “Medscape national physician burnout & suicide report,” *Medscape*.

Krijnen, Job M. T., David Tannenbaum, and Craig R. Fox (2017), “Choice architecture 2.0: Behavioral policy as an implicit social interaction,” *Behavioral Science & Policy*, 3 (2), i–18.

Lockwood, Penelope and Ziva Kunda (1997), “Superstars and me: Predicting the impact of role models on the self,” *Journal of Personality and Social Psychology*, 73 (1), 91–103.

Maslach, Christina, Wilmar B Schaufeli, and Michael P Leiter (2001), “Job burnout,” *Annual Review of Psychology*, 52 (1), 397–422.

Mayer, Thom, Arjun Venkatesh, and Donald M. Berwick (2021), “Criterion-based measurements of patient experience in health care,” *JAMA*, 326 (24), 2471.

McKethan, Aaron and Ashish K. Jha (2014), “Designing smarter pay-for-performance programs,” *JAMA*, 312 (24), 2617–18.

Meeker, Daniella, Jeffrey A. Linder, Craig R. Fox, Mark W. Friedberg, Stephen D. Persell, Noah J. Goldstein, Tara K. Knight, Joel W. Hay, and Jason N. Doctor (2016), “Effect of behavioral interventions on inappropriate antibiotic prescribing among primary care practices," *JAMA*, 315 (6), 562.

Navathe, Amol S., Kevin G. Volpp, Amelia M. Bond, Kristin A. Linn, Kristen L. Caldarella, Andrea B. Troxel, Jingsan Zhu, Lin Yang, Shireen E. Matloubieh, Elizabeth E. Drye, Susannah M. Bernheim, Emily Oshima Lee, Mark Mugiishi, Kimberly Takata Endo, Justin Yoshimoto, and Ezekiel J. Emanuel (2020), "Assessing the effectiveness of peer comparisons as a way to improve health care quality," *Health Affairs*, 39 (5), 852–61.

Ranganathan, Aruna and Alan Benson (2020), "A numbers game: Quantification of work, auto-gamification, and worker productivity," *American Sociological Review*, 85 (4), 573–609.

Richer, Sylvie F and Robert J Vallerand (1998), "Construction and validation of the ESAS (The relatedness feelings scale)," *European Review of Applied Psychology*, 48, 129–38.

Rogers, Todd and Avi Feller (2016), "Discouraged by peer excellence: Exposure to exemplary peer performance causes quitting," *Psychological Science*, 27 (3), 365–74.

Saccardo, Silvia, Hengchen Dai, Maria Han, Vangala Sitaram, Hoo Juyea, and Jeffrey Fujimoto (2023), "Field Tested : Assessing the Transferability of Behavioral Interventions."

Shanafelt, Tait D., Grace Gorringe, Ronald Menaker, Kristin A. Storz, David Reeves, Steven J. Buskirk, Jeff A. Sloan, and Stephen J. Swensen (2015), "Impact of organizational leadership on physician burnout and satisfaction," *Mayo Clinic Proceedings*, 90 (4), 432–40.

Song, Hummy, Anita L. Tucker, Karen L. Murrell, and David R. Vinsonc (2018), "Closing the productivity gap: Improving worker productivity through public relative performance feedback and validation of best practices," *Management Science*, 64 (6), 2628–49.

Sparkman, Gregg and Gregory M. Walton (2017), "Dynamic norms promote sustainable behavior, even if it is counternormative," *Psychological Science*, 28 (11), 1663–74.

Staw, Barry M., Robert I. Sutton, and Lisa H. Pelled (1994), “Employee positive emotion and favorable outcomes at the workplace,” *Organization Science*, 5 (1), 51–71.

Tran, Anh and Richard Zeckhauser (2012), “Rank as an inherent incentive: Evidence from a field experiment,” *Journal of Public Economics*, 96 (9–10), 645–50.

Verbeke, Willem, Richard P. Bagozzi, and Frank D. Belschak (2016), “The role of status and leadership style in sales contests: A natural field experiment,” *Journal of Business Research*, 69 (10), 4112–20.

Vidal, Jordi Blanes I. and Mareike Nossol (2011), “Tournaments without prizes: Evidence from personnel records,” *Management Science*, 57 (10), 1721–36.

Weick, Karl E., Kathleen M. Sutcliffe, and David Obstfeld (2005), “Organizing and the process of sensemaking,” *Organization Science*, 16 (4), 409–21.

West, C. P., L. N. Dyrbye, and T. D. Shanafelt (2018), “Physician burnout: contributors, consequences and solutions,” *Journal of Internal Medicine*, 283 (6), 516–29.

Wright, Thomas A. and Douglas G. Bonett (1997), “The contribution of burnout to work performance,” *Journal of Organizational Behavior*, 18 (5), 491–99.

Yates, Scott W. (2020), “Physician stress and burnout,” *American Journal of Medicine*, 133 (2), 160–64.

CHAPTER 3: WHEN IMPACT APPEALS BACKFIRE

EVIDENCE FROM A MULTINATIONAL FIELD EXPERIMENT AND THE LAB

Joseph S. Reiff [a]

Hengchen Dai [a]

Jana Gallus [a]

Anita McClough [b]

Steve Eitniear [b]

Michelle Slick

Charlotte Blank

[a] University of California, Los Angeles
[b] General Motors

Acknowledgment: We thank Russ Frey, InMoment, Maritz, and our partners at the technology company for their help implementing the field experiment. We thank the Customer Experience Professionals Association (CXPA) and the MBA marketing association we partnered with for surveying their members and sharing the data. We also thank Franklin Shaddy, Hal Hershfield, Megan Weber, and Stephanie Tjoa for their helpful comments, and we thank Jose Cervantez and Helen Zhang for their excellent research assistance.

Financial Disclosure: This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1650604. We acknowledge financial support from the Morrison Center for Marketing and Data Analytics, UCLA Anderson School of Management, UCLA Anderson Behavioral Lab, and the UCLA Academic Senate Council on Research Faculty Grants Program. The opinions and conclusions expressed are solely those of the authors and do not represent the opinions or policy of InMoment, Maritz, or the technology company we partnered with.

Abstract: Firms often try to motivate customers to share feedback by telling customers that their feedback will have an impact on the organization (e.g., “Have your say in our company’s direction”). We examine whether such “impact appeals” indeed increase compliance with customer feedback requests. In a field experiment across seven countries, 430,666 customers of a Fortune 500 company received a customer feedback survey invitation email where the subject lines were manipulated. Contrary to our initial prediction and expert forecasts, we found that impact appeals on average decreased feedback provision (compared to a straightforward control message). Importantly, impact appeals backfired to a greater extent in countries with lower trust in business (e.g., Japan) than in countries with higher trust in business (e.g., China). We theorize that impact appeals are more likely to reduce compliance among customers with lower trust in business because these customers perceive impact appeals as more inauthentic. Pre-registered lab experiments (N=7,926) support our theoretical account and test more effective impact appeals informed by it. This research sheds light on when and why highlighting impact can fail to motivate customers and even backfire, and more generally, it advances the field’s understanding of what drives customer engagement in empowering behaviors.

Keywords: customer feedback, empowerment, inauthenticity, trust, social influence, field experiment

INTRODUCTION AND THEORY

Soliciting customers' feedback is a common way that firms learn about the customer experience, improve their products and services, and keep customers engaged. Customer feedback helps firms identify barriers to customer satisfaction, improve product development, and conduct customer segmentation (Berry, Carbone, and Haeckel 2002; Lemon and Verhoef 2016; Meyer and Schwager 2011; Morgan, Anderson, and Mittal 2005). Beyond these direct benefits for firms, customers who contributed feedback typically develop more positive perceptions of the firm, purchase more, and exhibit greater brand loyalty (Bone et al. 2017; Chandon, Morwitz, and Reinartz 2004; Dholakia 2010; Dholakia and Morwitz 2002; Liu and Gal 2011; Morwitz, Johnson, and Schmittlein 1993). While previous research has identified these important downstream consequences of customer feedback provision, it is largely unknown how to effectively motivate customers to provide feedback in the first place.

To increase compliance with feedback requests, firms often tell customers that their feedback can impact the organization—a tactic that we call an "impact appeal" (e.g., "You can help shape the Delta experience", "Have your say in Prolific's direction"). In a survey we conducted with customer experience professionals (N = 76; Web Appendix A), we asked which strategies their organizations most frequently use to collect feedback: 48.7% of professionals reported using impact appeals, followed by 32.9% who reported using neutral requests that simply state the call to action (e.g., "Bank of America Requests Your Feedback"), and only 11.8% who reported offering rewards. Similarly, customers whom we surveyed (N = 201; Web Appendix A) on average reported that they "frequently" received customer feedback solicitations with impact appeals or neutral requests (and "rarely" received solicitations offering rewards).[18]

---

[18] Participants rated three strategies—impact appeals, neutral requests, and rewards—on a 1-6 scale: 1 – Never, 2 – Very Rarely, 3 – Rarely, 4 – Occasionally, 5 – Frequently, and 6 – Very Frequently. The median values were 5 for

Despite their widespread use, research has yet to investigate whether impact appeals are effective in motivating feedback provision.

We conducted a multinational natural field experiment and two controlled lab studies to examine the effect of impact appeals on customers' compliance with feedback requests. The field experiment shows that, contrary to both our initial prediction and the forecasts of marketing experts, impact appeals on average *decreased* customers' likelihood of offering feedback (compared to a neutral control request). Yet, this average effect masks important heterogeneity across countries. We develop and test a theoretical account about the roles of perceived inauthenticity and trust in business, which can explain this cross-cultural heterogeneity and, more generally, when impact appeals motivate feedback provision and when they backfire. By testing different customer feedback requests and examining the psychological processes behind customer responses, the current research simultaneously advances theoretical understanding of influence tactics while offering prescriptive implications for marketers who hope to motivate this important customer behavior.

THEORETICAL DEVELOPMENT

Self-efficacy and Empowerment

Impact appeals are an influence tactic in which firms attempt to motivate consumers to take an action by telling them that their action can impact the organization. We initially theorized that impact appeals would motivate compliance with feedback requests because people value having an impact. This prediction is consistent with the broad idea in psychology and marketing that people have an innate desire to perceive personal control over outcomes (Langer 1975;

---

both impact appeals and neutral requests versus 3 for rewards. To validate these judgments based on participants' recollection, participants also shared their most recently received email containing a customer feedback request. This data confirmed that reward offerings were less common than impact appeals.

Cutright and Wu 2023). More specifically, our initial hypothesis was inspired by the literatures on self-efficacy and customer empowerment.

Self-efficacy, or the belief that one can take steps to make an impact, is linked to heightened motivation across domains (see Bandura 2006 for a review). For instance, self-efficacy positively predicts work performance (Stajkovic and Luthans 1998), academic achievement (Multon, Brown, and Lent 1991), and healthy behaviors (Holden 1992). Relatedly, the belief that individual political action can change the government yields greater voter turnout (Abramson and Aldrich 1982), and the belief that one's donation is impactful boosts charitable giving (Cryder, Loewenstein, and Scheines 2013; Sharma and Morwitz 2016). As explained by Bandura (2001, p. 10), "Unless people believe they can produce desired results and forestall detrimental ones by their actions, they have little incentive to act or to persevere in the face of difficulties."

The empowerment literature further suggests that customers generally value having impact and that they engage more with firms that make them feel like they can shape the organization (Fuchs and Schreier 2011; Wathieu et al 2002). For example, customers have more positive brand attitudes and stronger purchasing intentions after they contribute to the creation of a firm's products (i.e., "co-creation"; Franke, Schreier, and Kaiser 2010; Füller et al. 2009) or after they share their opinions about products (Fuchs, Prandelli, and Schreier 2010). Such "empowerment strategies" are theorized to generate positive outcomes by increasing individuals' perceptions of their personal impact on the organization (Fuchs, Prandelli, and Schreier 2010; Spreitzer 1995).

The self-efficacy and customer empowerment literatures therefore led to our initial hypothesis, that by highlighting the prospect of having impact, impact appeals would motivate customers to offer feedback. To our knowledge, previous research has not tested whether prospectively framing an action as empowering—highlighting its potential impact on the organization—would motivate customers to engage in that action.

We tested this initial hypothesis in a large-scale field experiment with a Fortune 500 technology company (Study 1). Over 400,000 customers across seven countries around the world were invited by email to provide feedback on their recent customer service experience with the company. The emails differed in whether the subject line contained one of three variants of an impact appeal (e.g., "Your voice is important: Shape [Company]'s customer experience") or a control subject line ("[Company] customer experience survey invitation").[19] The primary outcome was the share of customers who started the survey in response to each subject line ("response rate").

Expert Predictions

To quantitatively assess marketing experts' beliefs about the effectiveness of impact appeals in this context, we conducted two pre-registered surveys. In one survey, we partnered with the Customer Experience Professionals Association (CXPA) to poll their members and followers. Respondents learned about the context and design of the field experiment, as well as the response rate among USA-based customers who received the control subject line. They then predicted USA-based customers' survey response rate for each impact appeal; The 76 experts who had worked in customer experience or marketing expected impact appeals to increase the

---

[19] The field experiment also tested a subject line that was identical to the control subject line but it also included the expected time it would take to complete the survey (i.e., "[Company] customer experience survey invitation (only takes 2 minutes)"). See Study 1, Part 1 for more information.

response rate relative to the control subject line (by an average 5.4 percentage points).[20] Thus, these marketing experts shared our initial hypothesis about the effectiveness of impact appeals.

The second survey involved 68 members of a large business school's marketing association, all of whom had worked in marketing, taken graduate-level courses in marketing, or both. Consistent with the results from the first survey, respondents were more likely to predict that impact appeals and not the control subject line would most effectively motivate customers to provide feedback. See Web Appendix A for detailed protocols and results for both expert surveys.

Preview of Field Experimental Results

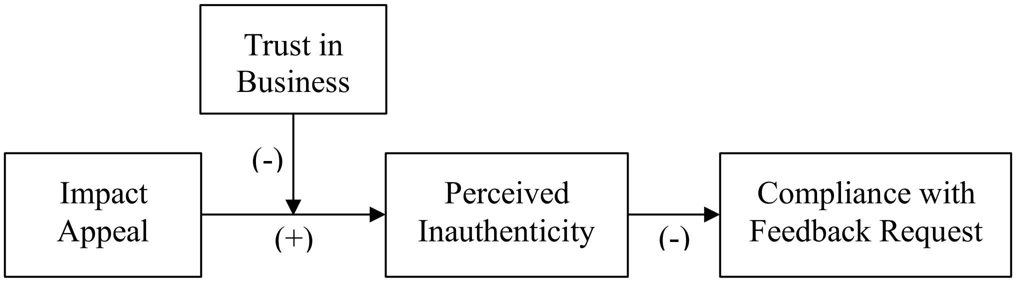
Contrary to our initial hypothesis and expert predictions, impact appeals on average *decreased* customers' willingness to provide feedback in the field experiment. Importantly, the effects of the impact appeals varied across countries: The impact appeals backfired in Japan, France, Brazil, and USA; they had no statistically significant effects in Canada and Germany; and they increased feedback provision in China. The negative main effect and substantial cross-country heterogeneity led us to develop and test a theoretical account about when and why impact appeals fail to motivate customers and even backfire.

We propose that impact appeals are more likely to reduce compliance with feedback requests among customers with lower trust in business because these customers perceive impact appeals as more inauthentic (Figure 3-1 summarizes the theoretical model). We next develop this theoretical account, drawing on literature in marketing, management, and psychology.

---

[20] Since a large proportion of CXPA's members and social media followers work in the USA, we designed the survey to collect predictions about the USA-based customers in our field experiment. Among the 42 marketing professionals who self-reported living in the USA, they similarly expected impact appeals to increase the response rate (by an average of 3.7 percentage points.)

Figure 3-1. Theoretical Model



The Limits of Persuasion

For an impact appeal to increase compliance with feedback requests, presumably customers must believe the appeal's claim that customer feedback could have the stated impact. Despite proposing verbal persuasion as a key source of self-efficacy, Bandura (1977) himself warned that such tactics may be too weak to instill efficacy expectations if people do not have a lived experience of generating impact in a given situation. In that sense, firms' attempts to use impact appeals may not yield any motivational effects as they could simply be seen as cheap talk.

Even worse, we propose that impact appeals may elicit negative perceptions about firms. Our theorizing is built on the psychology and marketing literatures suggesting that people are vigilant of firms' persuasion tactics. In particular, the Persuasion Knowledge Model (Friestad and Wright 1994) posits that people accumulate knowledge about marketers' persuasion tactics, which they use to recognize and cope with persuasion attempts. The Persuasion Knowledge Model further proposes that people draw inferences about marketers' intentions based on the tactics in use, which subsequently influence their decisions (Campbell and Kirmani 2000; Kirmani and Campbell 2004). For example, when consumers perceive that a firm is attempting to control their behavior and limit their freedom, they may exhibit reactance and avoid transacting with that organization in the future (Brehm 1966; Clee and Wicklund 1980; Fitzsimons and Lehman 2004). The use of attention-getting tactics (e.g., including a puppy in an advertisement for cookware; Campbell 1995) or statements derogating the competition (Jain and Posavac 2004) can lead people to infer that the firm is deploying a manipulative persuasion tactic, which lowers consumers' purchasing intentions. In the context of customer feedback, suspicions of marketers' persuasive intent have been found to attenuate the otherwise positive effects of answering marketers' survey questions (Williams, Fitzsimons, and Block 2004). Extending this literature, we theorize how and when consumers' inferential processes may lead them to perceive inauthenticity from marketers' use of impact appeals, which can influence their compliance with marketers' feedback requests.

Perceived Inauthenticity

We define perceived inauthenticity in the context of social influence as a negative judgment about an influence tactic that arises when consumers perceive a mismatch between the intentions a firm deliberately communicates (i.e., stated intentions) and the firm's presumed true intentions. Perceptions of inauthenticity can arise for multiple reasons (Silver, Newman, and Small, 2021): Consumers may perceive a firm's stated intentions to be entirely different from its true intentions (i.e., it has ulterior motives), or they may view a firm's stated intentions as an exaggeration of its true intentions. In our context, when a firm uses an impact appeal, customers may perceive a mismatch between the firm's stated intentions of using the feedback to impact the organization and the firm's true intentions. This mismatch could arise due to perceived ulterior motives if customers infer, for example, that the firm's true intention is to use impact appeals to promote its brand image by leaving the positive impression that it "listens" to its customers (DeCarlo 2005; Morrison and Bies 1991). Or customers may feel that the firm purposefully overstates its true intention about how it plans to use the feedback, exaggerating customers' potential impact (Darke and Ritchie 2007; Falbo 1977). These perceptions would lead customers to judge the firm's impact appeal as inauthentic.

When customers feel that a firm's influence tactic is an inauthentic representation of its true intentions, they typically become less willing to comply with the firm's request (Amengual and Apfelbaum 2021; Barasch, Berman, and Small 2016; Wang, Krishna, and McFerran 2017). In the context of feedback solicitation, if customers view the use of impact appeals as inauthentic, impact appeals may inadvertently reduce customers' willingness to provide feedback. Next, we theorize about *when* impact appeals are more likely to elicit perceptions of inauthenticity and reduce customers' compliance with feedback solicitation requests.

Trust in Business

We propose that customers' level of trust in business play a pivotal role in shaping their perceptions and subsequent responses to impact appeals. Generally speaking, when customers do not trust organizations, they may be more likely to scrutinize firms' marketing messages (Priester and Petty 1995). In this sense, consumers' desire to understand marketers' underlying intentions—a key component of the Persuasion Knowledge Model (Friestad and Wright 1994)—is enhanced when consumers have lower trust in business (Kirmani and Campbell 2004). Moreover, upon scrutinizing marketers' messages, customers with lower trust in business are also more skeptical about the nature of marketers' underlying intentions and more prone to make negative inferences about them (Chaudhuri and Holbrook 2001; Mayer, Davis, and Schoorman 1995; Kirmani and Campbell 2004; Rajavi, Kushwaha, and Steenkamp 2019). For instance, when customers are less trusting of a firm, they react more negatively to strategies that make expensive options pre-selected defaults (Brown and Krishna 2004). In advertising, customers who have been targeted by deceptive advertisements subsequently become less trusting of advertisements in general, causing them to evaluate future advertisements as more deceptive (Darke and Ritchie 2007). Relatedly, when firms have worse reputations, their extreme claims are more strongly discounted (Goldberg and Hartwick 1990).

Building on this research, we theorize that customers' trust in business influences how they perceive the intentions behind marketers' use of impact appeals. Since customers with lower trust in business apply greater scrutiny and skepticism towards firms' tactics and underlying intentions, these customers are more likely to infer from a firm's use of an impact appeal that the firm has ulterior motives, that the firm is exaggerating how much it actually plans to rely on the customer feedback, or both. Customers with lower trust in business may thus perceive impact appeals as more inauthentic. As we previously theorized, such perceptions of inauthenticity should reduce customers' willingness to comply with the feedback requests.

Bringing this all together, we propose three hypotheses. We first predict that customers with lower trust in business will perceive impact appeals as more inauthentic, and that perceptions of inauthenticity will negatively predict compliance with feedback requests.

> ***Hypothesis 1a**: Impact appeals increase perceptions of inauthenticity to a greater extent when customers have lower (vs. higher) trust in business.*
>
> ***Hypothesis 1b**. Customers with stronger perceptions of inauthenticity are less willing to provide feedback.*

Given Hypotheses 1a and 1b, we can predict when impact appeals will be more likely to backfire.

> ***Hypothesis 2**: Impact appeals are more likely to reduce feedback provision when customers have lower (vs. higher) trust in business.*

To summarize, our model posits that impact appeals are more likely to reduce feedback provision among customers with lower trust in business because these customers perceive impact appeals as more inauthentic (see Figure 3-1).

Together, our research makes five key theoretical contributions. First, while prior research has explored the *downstream consequences* of customer feedback provision (Bone et al. 2017; Chandon, Morwitz, and Reinartz 2004; Dholakia 2010; Dholakia and Morwitz 2002; Kim et al. 2019; Lemon and Verhoef 2016; Morwitz, Johnson, and Schmittlein 1993), we examine the *antecedents* of customer feedback provision—that is, what affects the adoption of this important customer behavior. Second, extending prior research suggesting that customers value feeling empowered (Fuchs and Schreier 2011; Fuchs, Prandelli, and Schreier 2010; Wathieu et al. 2002), we study the effects of prospectively *framing* an action as empowering—by highlighting its potential impact on the organization. Third, while previous research suggests that it can be motivating to tell people that their acts are consequential (e.g., Cryder, Loewenstein, and Scheines 2013; Sharma and Morwitz 2016), we shed light on when and why appealing to impact reduces customer engagement. Fourth, we contribute to work on perceptions of inauthenticity by showing that this construct may help explain the effects of impact appeals, and we empirically delineate two sources of inauthenticity recently proposed by Silver, Newman, and Small (2021). Fifth, in response to a call for more cross-cultural research on the Persuasion Knowledge Model (Campbell and Kirmani 2008), we provide some of the first field evidence showing that responses to the same influence tactic may differ widely across cultures, and we propose trust in business as a moderator that can help explain this cross-cultural heterogeneity. Overall, the current research provides a “phenomenon-to-construct mapping” by identifying and helping explain an otherwise puzzling real-world phenomenon—the unexpected negative effects of a commonly used marketing tactic (Lynch and van Osselaer 2022).

We focus on reporting one field and two lab experiments and briefly summarize insights from three additional experiments described in detail in the Web Appendix. We first present the design and results of the field experiment that tested the effects of impact appeals on customers' likelihood of providing feedback across seven countries (Study 1). We report the field experiment in two parts: In part 1, we report results from our pre-registered analyses, which reveal a *negative* average treatment effect of impact appeals, and in part 2, we analyze the data to test our theoretical account about when impact appeals backfire. We then report two pre-registered lab experiments that deductively test all three hypotheses (Studies 2 and 3).

Pre-registrations, study materials, non-proprietary data, and code are available at https://researchbox.org/433&PEER_REVIEW_passcode=ELFWGO. All analytic choices in this paper follow pre-registrations unless otherwise specified.

STUDY 1, PART 1: IMPACT APPEALS ACROSS THE WORLD

Methods

Setting

We collaborated with a Fortune 500 technology company that sells computer hardware and software. Like many other firms, the company routinely sends customers an email inviting them to complete a voluntary customer feedback survey approximately 24 hours after they receive customer service help. The survey remains open for 10 days following the initial invitation date. If a customer does not respond to the initial request within 72 hours, they receive an identical invitation as a reminder. We conducted a pre-registered field experiment in this context.

Experimental Design

The experiment ran for eight weeks (from 10/12/2020 through 12/06/2020) and involved all customers in Brazil, Canada, China, France, Germany, Japan, and the USA who received customer service help from the company during this period and who used the respective country's primary language in their communication with the firm.[21] An a priori power calculation suggested that running the experiment for eight weeks would give us 80% power to detect a 10% relative increase[22] in the share of customers who start the survey within most countries in the study (and that we would be slightly underpowered to detect this effect size within Canada, France, and Germany). The relative effect size of 10% is a conservative lower bound based on the effects observed in another field experiment involving an email campaign (Sahni, Wheeler, and Chintagunta 2018).

During the experiment, customers were randomly assigned to one of five conditions at an equal probability prior to receiving their initial survey invitation. We exogenously varied the email subject lines of these invitations. The subject line either straightforwardly invited customers to take the survey (Control; the company's status quo), additionally indicated the short amount of time the survey was estimated to take (Time), or contained an impact appeal (Impact). We had three variants of impact appeals: Impact (Voice), Impact (Help), and Impact (Expert). Specifically, the five subject lines read:

Control: *[Company] customer experience survey invitation*
Time: *[Company] customer experience survey invitation (only takes 2 minutes)*
Impact (Voice): *Your voice is important: Shape the [Company] customer experience*
Impact (Help): *Your help is needed: Shape the [Company] customer experience*
Impact (Expert): *Your expert advice is appreciated: Shape the [Company] customer experience*

---

[21] Customers in Canada received the same English subject lines as customers in the USA.

[22] This effect size is relative to the share of customers who started the survey in response to the control subject line (i.e., the status quo) within each country during the month prior to the experiment (April 2020).

The Time subject line was included as a benchmark, to see whether the impact appeals could outperform a message that highlighted the low time costs of taking the survey. The experiment included three variants of impact appeals for stimuli sampling purposes—to ensure the effects would not be driven by any one specific instantiation of an impact appeal. This approach also has the advantage that it reflects practice; there are different ways that firms can convey customers can make an impact—for instance, customers can make an impact by offering their help or expertise—and thus it was important to test different instantiations of impact appeals. The subject lines were professionally translated into the respective native language (see Table A3-2 in Web Appendix B for the translations).

The email subject lines were pre-tested on Prolific, where we recruited bi-lingual speakers (N = 954) who were fluent in English and one of the other five languages. We confirmed that, across all languages, participants rated the impact appeals as more strongly conveying that customer feedback would have an impact, compared to the control subject line (*p*-values < .01; see Web Appendix C for more information about the pre-test).

The email body was identical across conditions and used a standard customer feedback invitation format. Specifically, the email body briefly explained that the feedback would be useful for the company. It then displayed a standard customer satisfaction question. If customers clicked on the question, they were directed to the survey, which contained three questions about their recent customer experience and took customers approximately 80 seconds to complete (median).

Sample Characteristics and Balance Checks

As pre-registered, we focused the analysis on the first instance when a customer received service help during the 8-week experimental period. We confirmed that email bounce rates were balanced across conditions (p = .757 for the F-test of joint significance) and thus excluded customers whose email invitation bounced, following our pre-registration. The final sample includes 430,666 customers. Table 3-1 displays the sample size by country and language.

Table 3-1. Sample Size by Country in Study 1.

Country	Language	N
Brazil	Brazilian Portuguese	28,559
Canada	English	13,926
China	Simplified Chinese	137,890
France	French	16,769
Germany	German	17,603
Japan	Japanese	36,435
USA	English	179,484
		Total = 430,666

At the time of data collection, 91.2% of the final sample had a warranty contract, 48.8% had a premium account, 47.8% had purchased their computer for personal use (rather than for business or other non-personal use), and 50.8% were customers of the company's subsidiary firm. 62.7% had not received a customer feedback survey invitation in the previous year, and 16.7% had received just one invitation. With respect to the specific service they were asked to provide feedback on during our study period, 77.5% of customers had received customer service help remotely (either online or over the phone) rather than in-person. Demographic information about the sample was not available. Table A3-5 (Web Appendix D) shows results from a balance test, which confirms that customers do not significantly differ between conditions in the aforementioned baseline characteristics.

Outcome Variables and Analysis Strategy

Our primary outcome of interest, *survey starting,* is a binary variable indicating whether a customer started the survey within 10 days of receiving the invitation (i.e., before the survey expired). Hereafter, we refer to the share of customers who started the survey as the response rate. We report two pre-registered secondary outcomes: *survey completion*, measured with a binary variable indicating whether a customer completed the survey within 10 days of receiving the invitation (i.e., before the survey expired), and *unsubscribing,* a binary variable that indicates whether, during the 8-week experimental period, the customer unsubscribed from future customer feedback invitation emails. We were unable to track email opening due to technological restrictions.

To test the overall effect of impact appeals, our primary pre-registered analysis looks at the combined effect of the three variants of impact appeals. The main model is an OLS regression with the outcome variable on the left-hand side, where the right-hand side includes a dummy variable (*Impact*) indicating whether participants were in one of the Impact conditions and another dummy variable (*Time*) indicating whether participants were in the Time condition. The Control condition is the reference group. The model includes language and week fixed effects and uses heteroskedasticity robust standard errors. The model relies on intention-to-treat and thus includes the full sample of customers (N = 430,666).

To study the effects of each experimental condition, we also estimated a pre-registered secondary model that is identical to the main model except that it includes a separate dummy variable for each of the treatment conditions (rather than grouping the three Impact conditions together). To estimate treatment effects within each country, we used separate OLS regressions, each containing one country's data. The right-hand side of the regressions includes the *Impact* dummy variable, the *Time* dummy variable, and week fixed effects, with heteroskedasticity robust standard errors.

Results

Main Effects Aggregated Across Countries

Our pre-registered hypothesis was that the impact appeals would increase customers' likelihood of sharing their feedback with the company. Yet, in contrast to both our initial hypothesis and expert predictions (described earlier), the Impact conditions resulted in a lower average response rate (25.7%) than the Control condition (26.1%). Using our main model, we estimated that the impact appeals significantly reduced the likelihood of survey starting by .4 percentage points (b = -.004, p = .015), or a 1.6% relative decrease compared to the Control condition.[23] The secondary analysis shows that all three impact appeals either statistically significantly decreased survey starting or had an effect statistically indistinguishable from zero: Impact (Voice), b = -.010, p < .001; Impact (Help), b = .002, p = .254; Impact (Expert), b = -.005, p = .031. The Time condition did not have a statistically significant effect on survey starting (b = .003, p = .183). The regression tables for the primary and secondary analyses are reported in Web Appendices E and F.

Looking at survey completions, we estimated that the impact appeals significantly reduced the likelihood of survey completion by .9 percentage points (b = -.009, p < .001), representing a 4.1% decrease, compared to the 21.2% completion rate in the Control condition. All three impact appeals either significantly or marginally significantly decreased survey completion: Impact (Voice), b = -.013, p < .001; Impact (Help), b = -.003, p = .085; Impact (Expert), b = -.010, p < .001.

---

[23] Hereafter, we use "survey starting" for brevity when referring to the likelihood of survey starting.

We further estimated the effect of impact appeals on unsubscribing. Compared to the unsubscribe rate of 0.16% in the Control condition, the impact appeals significantly increased the likelihood that customers unsubscribed by 0.04 percentage points (b = .0004, p = .020), representing a 25% increase.[24] In other words, customers who received an impact appeal were more likely to unsubscribe in order to spare themselves from future communications, compared to customers in the Control condition.[25]

The results are robust to including the baseline sample characteristics as covariates (see Web Appendix E), as well as to using logistic regressions and mixed effects regressions with language and week random effects (see Web Appendix F). For the remaining analyses reported in the manuscript, we continue to follow our pre-registered plan to focus our analyses on the combined effect of the impact appeals.

Country-Specific Effects

Going beyond the negative average treatment effect whereby impact appeals reduced survey starting, we also uncovered considerable heterogeneity across countries. The impact appeals *reduced* survey starting in Japan (-9.2% change relative to the Control condition; b = -.042, p < .001), France (-5.8%; b = -.022, p = .022), Brazil (-5.6%; b = -.018, p = .011), and USA (-5.8%; b = -.015, p < .001); they did not have a statistically significant effect in Canada (-.2%; b = -.001, p = .957) and Germany (+4.1%; b = .015, p = .109); but the impact appeals significantly

---

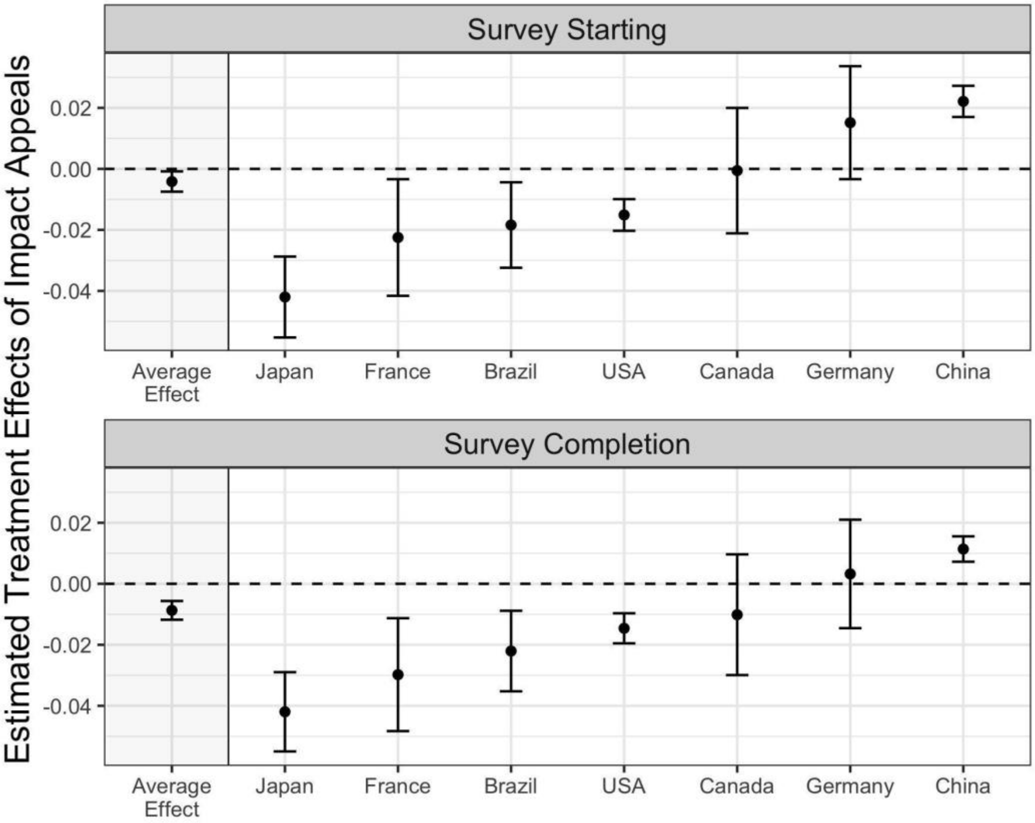
[24] The effect of the impact appeals on unsubscribing was primarily driven by the Impact (Help) condition (see Table A3-8 in Web Appendix F).

[25] We also examined whether satisfaction scores reported in the survey differed between conditions. Satisfaction scores could range from 0 (= Not at all Satisfied) to 10 (= Extremely Satisfied). Among the 108,671 customers who provided satisfaction scores, there was no statistically significant difference in scores between the Impact conditions (Mean = 8.25, SD = 3.01) and the Control condition (Mean = 8.26, SD = 3.00; b = -.02, p = .415). That said, the analysis of satisfaction scores naturally only includes people who took the survey. It is possible that the type of person who took the survey in the Impact conditions differed from the type of person who took the survey in the Control condition. This means that the satisfaction results could be partially driven by selection effects induced by the treatments.

increased survey starting in China (+13.5%; b = .022, *p* < .001). We observed similar heterogeneity in terms of survey completion. Figure 3-2 displays the average estimated treatment effects on survey starting and survey completion as well as the estimates by country.

Figure 3-2. Estimated Average and Heterogeneous Treatment Effects of Impact Appeals on Survey Starting and Survey Completion (Study 1).

The black dots reflect estimated treatment effects of the impact appeals (relative to Control), and the vertical black lines reflect the treatment effects' 95% confidence intervals. The far-left side of the figure shows the average treatment effects across all seven countries in the sample (using the main model), and the rest of the figure shows treatment effects within each country (using the country-specific models). The top panel shows the effects on survey starting and the bottom panel shows the effects on survey completion.



Discussion

Thus far, we have shown that impact appeals on average reduced survey starting, reduced survey completion, and increased unsubscribing. However, there was significant heterogeneity across countries. Why do impact appeals decrease customer engagement in some countries, but increase engagement in others? For customers who generally do not trust business, the impact appeal may feel like an inauthentic influence tactic. Perceiving this inauthenticity may reduce customers' willingness to comply with the firm's request. However, customers who generally trust business may not perceive the impact appeal as inauthentic; instead, they may be positively motivated by the impact appeal, as we initially theorized.

STUDY 1, PART 2: MODERATION BY TRUST IN BUSINESS

To test this theoretical account, we examined whether a country's level of trust in business moderated the effect of the impact appeals on feedback provision in the field experiment (Hypothesis 2). This hypothesis was developed after we had run the field experiment, and therefore, this analysis was not pre-registered.

Methods

Trust in Business Measure

We merged the data from the field experiment with country-level data about trust in business from the Edelman Trust Barometer (Edelman 2021). Edelman is a global communications firm that regularly conducts surveys around the world assessing people's opinions about various types of organizations. We were able to draw on survey data that overlapped with our experimental period. Specifically, Edelman conducted its annual survey from October 19, 2020 to November 18, 2020, across 28 countries, including all the countries in our field experiment. The survey included approximately 1,150 respondents in each country, except for the USA and China, which had 1,650 respondents each. Edelman used nationally representative sampling on age, gender, education, and region (Edelman 2021).

The survey asked respondents, “Below is a list of institutions. For each one, please indicate how much you trust that institution to do what is right.” The list included “Government in general”, “NGOs in general”, “Media in general”, and “Business in general.” Participants rated their degree of trust in each type of institution on a 9-point scale (1 - Do not trust them at all; 9 – Trust them a great deal). To assess country-level trust in business, which is our measure of interest, Edelman calculated the percent of people within each country that selected one of the top 4 scale points and thus indicated some degree of trust in business in general. For instance, China’s trust in business score of 70 reflects that 70% of respondents in China selected one of the top 4 scale points. This method of coding trust mitigates cultural differences in Likert scale response styles (Chen, Lee, and Stevenson 1995). For instance, western cultures typically use more extreme scale points than eastern cultures, which makes it difficult to interpret mean differences across countries (i.e., differences could reflect true variation in a psychological construct or artifactual variation based on how the scale points were used by respondents). Nevertheless, we obtained similar results when using mean responses instead of Edelman’s coding scheme (see Web Appendix G).

Analysis Strategy

To test Hypothesis 2—that trust in business moderates the effect of impact appeals on feedback provision—we ran an OLS regression with the survey starting outcome on the left-hand side. The right-hand side includes the (mean-centered) country-level measure of trust in business, the *Impact* and *Time* dummy variables, and two interaction terms (between trust in business and each dummy variable). The model also includes week and country fixed effects, with standard errors clustered by country.

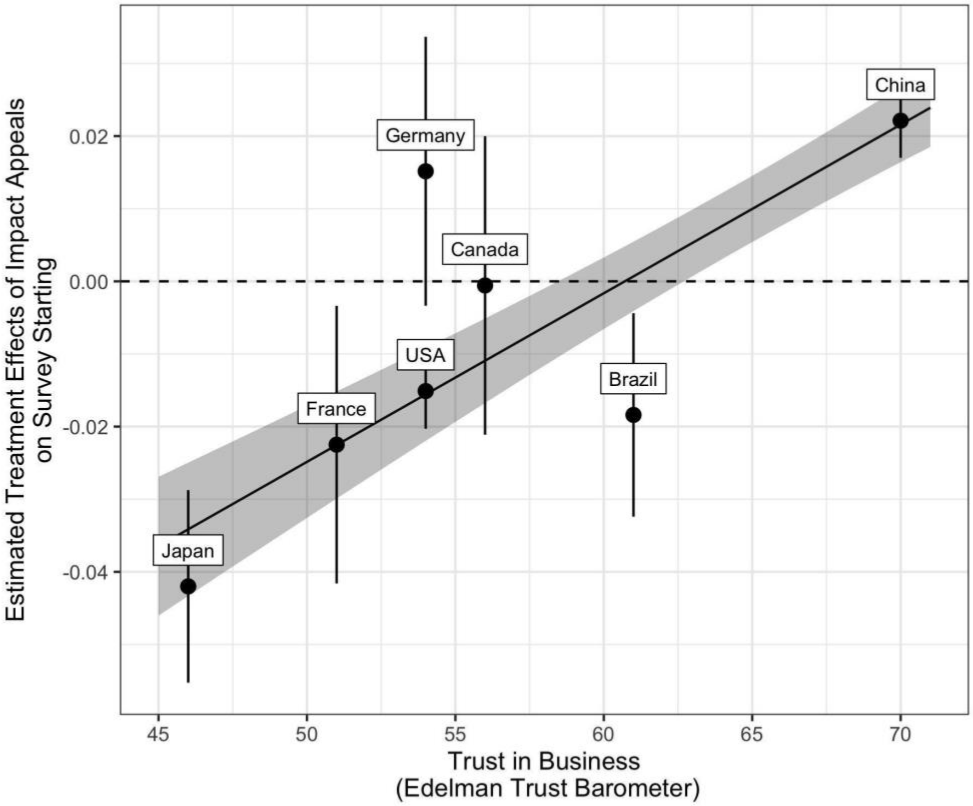
Results

Consistent with our theoretical account, the significant positive interaction between the impact appeals and trust in business (b = .002, $p < .001$) suggests that the impact appeals (vs. Control condition) had more negative effects on survey starting in countries with lower trust in business (e.g., Japan) than in countries with higher trust in business (e.g., China). That is, a 10-percentage-point decrease in the share of people who trust business within a country is associated with a 2-percentage-point decrease in the estimated treatment effect of impact appeals (relative to the control condition) on survey starting. See Web Appendix G for the corresponding regression table.

Using the same regression model, we estimated the simple effects of the impact appeals on survey starting at each level of trust in business (plotted in Figure 3-3). For example, for countries whose trust in business score was 45, close to Japan's level of trust, the impact appeals were estimated to decrease survey starting by 3.6 percentage points (b = -.036, $p < .001$) compared to the Control group; for countries whose trust in business score was 70, close to China's level of trust, the impact appeals were estimated to *increase* survey completion by 2.2 percentage points (b = .022, $p < .001$). Figure 3-3 also displays the regression-estimated treatment effect of impact appeals within each country (reported in Part 1). The relationship between the regression-estimated treatment effects and trust in business across countries depicted in Figure 3-3 reflects a correlation of Pearson $r = .726$.

Figure 3-3. Impact Appeals' Simple Effects and Estimated Treatment Effects by Trust in Business (Study 1).

Using the moderation regression reported in Part 2, the black slope represents the regression-estimated simple effects of the impact appeals on survey starting at every level of trust in business across the range of countries in our sample. The gray band contains the simple effects' 95% confidence intervals. Using the country-specific OLS regressions reported in Part 1, the black dots reflect each country's estimated treatment effect of the impact appeals (relative to Control), and the vertical black lines reflect the treatment effects' 95% confidence intervals.



Similar moderation results were found in our analyses of the secondary dependent variables (see Web Appendix G for the corresponding regression tables). In countries with lower (vs. higher) trust in business, impact appeals (vs. Control condition) had more negative effects on survey completion (interaction b = .002, $p < .001$) and increased unsubscribing to a greater extent (interaction b = -.00003, $p = .003$).[26] These patterns add robustness to our primary analyses, which looked at survey starting. The unsubscribing dependent variable moreover lends support to our theoretical account: The impact appeals appear to have been received poorly by customers with lower trust in business, so much so that these customers were more willing to unsubscribe from future communications with the firm.

We conducted several additional robustness checks to test alternative explanations (all reported in Web Appendix G). First, to ensure the results were not driven by trust ratings that may have been affected by the COVID-19 pandemic (the Edelman survey was collected in November 2020), we conducted a robustness check using a measure of trust in business from 2019 (Edelman Trust Barometer 2020) and found consistent results. Second, the results were robust to re-estimating the moderation models with a different coding scheme for trust in business: using each country's mean trust rating, rather than using Edelman's recommended approach to capture the share of people within each country who trust business. Third, we found that trust in business remains a significant moderator when also controlling for relevant alternative moderators—including uncertainty avoidance, power distance, and individualism (Hofstede, Hofstede, and Minkov 2010).

Discussion

In support of Hypothesis 2, impact appeals decreased survey starting to a greater extent in countries with lower trust in business. While the field experiment was particularly informative due to its large sample size and naturally assessed behaviors, it has limitations: (1) we were

---

[26] For a 10-percentage-point decrease in the share of people who trust business within a country, the treatment effect of impact appeals (relative to the control condition) on unsubscribing is estimated to increase by 0.03 percentage points, which corresponds to a roughly 20% relative change compared to the unsubscribing rate in the control condition (i.e., 0.16 percentage points).

unable to directly measure perceptions of the appeals, so we could not assess the underlying mechanisms; (2) the moderator of trust in business was measured at the country level, so variation in this measure only came from seven countries; (3) we used a one-item scale to measure trust; and (4) we developed the current hypothesis about trust in business after analyzing our original pre-registered hypotheses. We therefore designed a series of pre-registered lab experiments to deductively test our full theoretical account while addressing these concerns.

STUDY 2: THE ROLE OF PERCEIVED INAUTHENTICITY

Study 2 uses a hypothetical vignette asking people how they would feel and respond if they were to receive a customer feedback survey invitation from their computer company (as in the field experiment). We randomized whether they saw a subject line containing an impact appeal or a control subject line. The study has four main objectives. First, we sought to conceptually replicate the findings from the field experiment and test Hypothesis 2 again—that the effect of impact appeals on compliance with feedback requests is moderated by trust in business. But this time we recruited people from a single country (USA) and used a multi-item, individual-level measurement of trust in business. Second, we investigated underlying mechanisms by testing if people with lower trust in business perceive impact appeals as more inauthentic (Hypothesis 1a). Third, we examined whether perceptions of inauthenticity predict intentions to take customer feedback surveys (Hypothesis 1b). Fourth, to provide a deeper understanding of our key mechanism, we explored how two common forms of inauthenticity—perceptions of ulterior motives and exaggerated intentions (Silver, Newman, and Small 2021)—explain the effects of impact appeals. Together, Study 2 tests our full model (Figure 3-1) that impact appeals are more likely to reduce feedback provision among customers with lower trust in business because these customers perceive impact appeals as more inauthentic.

Methods

We recruited participants on Amazon's Mechanical Turk who lived in the USA. Following our pre-registration, participants were only invited to take the survey if they passed a brief attention check and were using a brand-name computer (rather than having built their own computer). The final sample included 1,998 people (mean age = 41.3, SD age = 12.9; 56.2% female).

Participants were first asked what computer brand they were using. The response options included 11 of the most used computer brands and an "Other" option that allowed them to write-in the brand (done by 4.5% of participants). Participants then were asked to imagine that they received help from their given computer company's customer service the previous day, the customer service was satisfactory, and all their questions were promptly answered.[27] Next, participants imagined receiving an invitation from their computer company for a customer feedback survey that would take about 2 minutes. We randomly assigned participants to read either the control subject line ("[Company] customer experience survey invitation") or an impact appeal subject line ("Your voice is important: Shape the [Company] customer experience"). Both were identical to the subject lines used in the field experiment. Every participant's specific computer brand name was piped into the subject line in the place of "[Company]". We focused on the Impact (Voice) subject line from the field experiment because the other two appeals additionally incorporated complementary psychological constructs like prosociality and status.

Following the manipulation, participants rated the extent to which the company's use of the subject line seemed inauthentic. We used two items to measure *perceived inauthenticity*,

---

[27] Positive customer service experiences seem more common in our data than negative experiences. For example, the vast majority of rated customer service encounters with our field partner were positive (median = 10 (out of a maximum of 10), with nearly 80% of responses rated at or above 8).

adapted from a measure of perceived manipulative intent in Campbell (1995): “The way this subject line tries to persuade customers to take the survey seems insincere to me” and “[Company] tries to influence customers with this subject line in an inauthentic way” (1 - Strongly Disagree; 6 - Strongly Agree). The two items were highly correlated (Pearson r = .73) so we averaged them together to create a composite score of perceived inauthenticity. On the following screen, we asked participants about their *survey taking intentions*: “If you were in this situation, how likely would you be to take the customer experience survey?” (1- Extremely Unlikely; 6 - Extremely Likely).

Participants next completed scales measuring two sources of inauthenticity: perceptions of ulterior motives and exaggerated intentions. We used two items to measure perceptions of ulterior motives: “[Company] is using this subject line to look good to its customers” and “I think there is another reason why [Company] is using this subject line, beyond just seeking customer feedback” (1 - Strongly Disagree; 6 - Strongly Agree). The two items were positively correlated (Pearson r = .52) so we averaged them to create a composite score of *ulterior motives*. We used two items to measure perceived exaggeration: “[Company] is intentionally exaggerating the impact customers could have on its customer service experience by taking the feedback survey” and “[Company] is intentionally exaggerating customers’ ability to influence its customer service by taking the feedback survey” (1 - Strongly Disagree; 6 - Strongly Agree). The two items were highly correlated (Pearson r = .91) so we averaged them to create a composite score of *exaggeration*.

Participants then completed a number of scales for alternative mediators including: perceptions of their potential impact on the company (*perceived impact;* two items, r = .87; Spreitzer 1995; Fuchs, Prandelli, and Schreier 2010), perceptions that the company is controlling their behavior (*perceived threat to freedom,* two items, r = .76; Irmak, Murdock, and Kanuri 2020), perceptions that the company appreciates feedback (*feedback appreciation,* two items, r = 0.88), perceptions of the personal benefits they would receive from offering feedback (*personal benefits;* one item), and perceptions of the personal costs they would incur from offering feedback (*personal costs;* one item; See Web Appendix H for the scale items). After reporting their age and gender, participants indicated how often they had responded to customer feedback requests in the prior year.[28] They then responded to a four-item scale measuring their general trusting attitudes towards business, adapted from a measure of general interpersonal trust developed by Levine et al. (2018). Example items include “I believe that companies do not intentionally misrepresent their products and services to customers” and “If I purchased an item from a company and it did not work as advertised, I believe the company would offer me a full refund for the item” (1 - Strongly Disagree; 6 - Strongly Agree). The four items were averaged into a composite measure of *trust in business* (α = .83). See Web Appendix I for the full set of items. Participants also responded to a four-item scale assessing trust in their specific computer company (also adapted from Levine et al. 2018; see Web Appendix J; α = .84). The two sets of trust measures were counterbalanced in order. We report the results for trust in business in the manuscript and trust in the specific company in Web Appendix J; trust in the specific company shows similar results, though weaker in terms of statistical significance.

Results

---

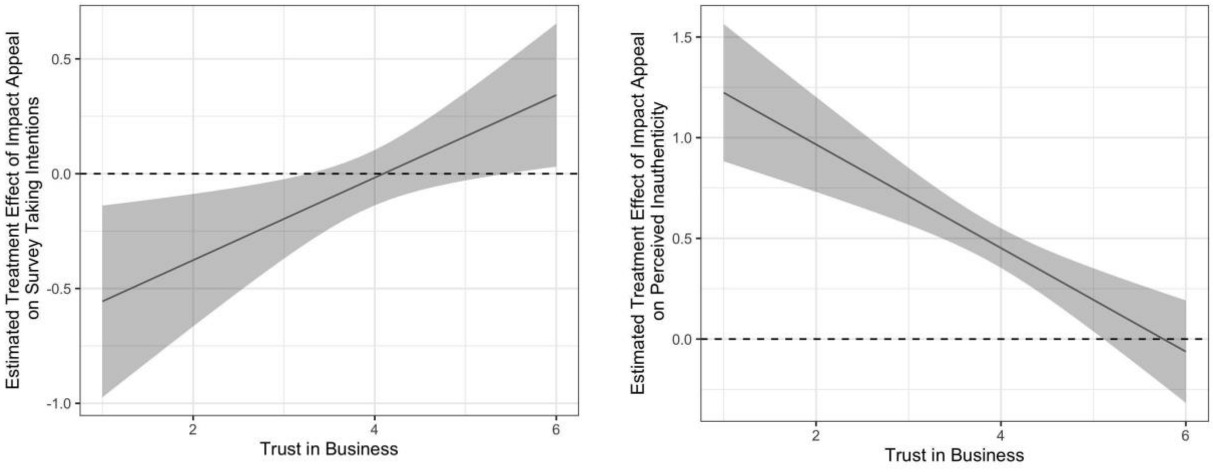
[28] We had originally pre-registered to include prior feedback provision as a covariate in all analyses for the purpose of increasing our statistical power. Only after collecting the data, we realized prior feedback provision could be causally affected by trust in business (whereby positive trusting attitudes towards businesses may lead customers to be willing to provide voluntary feedback); thus, it is not appropriate to include prior feedback provision as a covariate in the regression models that test the effect of trust in business (Angrist and Pischke 2008). We report results in the manuscript without controlling for prior feedback provision, though these results hold if we do control for it (Web Appendix I).

We first focus on Hypothesis 2 and test whether trust in business moderated the effect of the impact appeal on survey taking intentions. Before conducting our primary analysis, we confirmed that being randomly assigned to the Impact condition (vs. Control condition) did not affect responses to the trust in business moderator ($p = .665$). We then ran an OLS regression that predicts participants' survey taking intentions, where the right-hand side includes a dummy variable indicating whether participants were in the Impact (vs. Control) condition, (mean-centered) trust in business, and the interaction between the two variables. In line with our field study and Hypothesis 2, we found that the effect of the impact appeal on survey taking intentions depended on trust in business (interaction b = .180, $p = .011$; see Web Appendix I for the regression table). Figure 3-4 Panel A depicts the regression-estimated simple effects on survey taking intentions at every level of trust in business. When trust in business was low, the impact appeal had a negative effect on survey taking intentions. For example, for participants with moderately low trust in business (who tended to "disagree" with the trust in business items, with an average rating of 2 out of 6), the impact appeal was estimated to significantly reduce survey taking intentions (b = -.376, $p = .011$). However, this negative effect was attenuated as trust in business increased. For example, for people with moderately high trust in business (who tended to "agree" with the trust in business items, with an average rating of 5 out of 6), the impact appeal was estimated to increase survey taking intentions, albeit only marginally significantly (b = .163, $p = .097$). Together these results support Hypothesis 2 that the effect of impact appeals on survey taking intentions was more negative among individuals with lower trust in business.

We next tested whether trust in business moderated the effect of the impact appeal on perceptions of inauthenticity (Hypothesis 1a). We ran an OLS regression to predict perceived inauthenticity as a function of an indicator for the Impact (vs. Control) condition, (mean-centered) trust in business, and their interaction. See Web Appendix I for the regression table. The positive coefficient on the Impact indicator shows that compared to the Control condition, the Impact condition significantly increased perceived inauthenticity at the mean level of trust in business (b = .475, $p < .001$). Importantly, in support of Hypothesis 1a, the impact appeal increased perceived inauthenticity to a greater extent among people with lower trust in business (interaction b = -.257, $p < .001$). Panel B of Figure 3-4 displays the simple effects of the impact appeal on perceived inauthenticity at every level of trust in business. For people with moderately low trust in business (who on average indicated "disagree" on the trust in business items), the impact appeal was estimated to significantly increase perceived inauthenticity by 0.97 on a 6-point scale (b = .966, $p < .001$). The effect of the impact appeal on perceived inauthenticity was reduced to 0.20 for people with moderately high trust in business (who on average indicated "agree" on the trust in business items; b = .195, $p = .015$).

We further tested whether perceived inauthenticity helped explain the more negative effects of the impact appeal on survey taking intentions among people with lower trust in business. In support of Hypothesis 1b, we first confirmed that perceived inauthenticity negatively predicted survey taking intentions (b = -.381, $p < .001$). Next, we used moderated mediation analyses to test our full theoretical model. Specifically, we tested whether the indirect effect of the impact appeal on survey taking intentions through perceived inauthenticity depended on trust in business. Using 5,000 bootstrapped samples, we estimated that the impact appeal had a negative *indirect* effect on survey taking intentions via perceived inauthenticity among people with moderately low trust in business (who on average indicated "disagree" on the trust in business items; indirect effect = -.314, 95% CI = [-.432, -.218]). Yet for people with moderately high trust in business (who on average indicated "agree" on the trust in business items), the indirect effect of the impact appeals through perceived inauthenticity was significantly weaker (indirect effect = -.063, 95% CI = [-.121, -.007]; comparison of indirect effects between the two levels of trust, $p < .001$). These results suggest that perceptions of inauthenticity could help explain why the impact appeal decreased survey taking intentions to a greater extent among people with lower trust in business.

Figure 3-4. Treatment Effects of Impact Appeals Depend on Trust in Business (Study 2). This figure shows the regression-estimated simple effects of the impact appeal on survey taking intentions (Panel A) and perceived inauthenticity (Panel B) at every level of trust in business. The gray bands contain the 95% confidence intervals. The trust in business scale points correspond to the following labels: 1 = Strongly Disagree, 2 = Disagree, 3 = Somewhat Disagree, 4 = Somewhat Agree, 5 = Agree, and 6 = Strongly Agree.



Panel A. Survey Taking Intentions

Panel B. Perceived Inauthenticity

As a secondary analysis, we examined how the effects of impact appeals were explained by two common sources of inauthenticity: perceptions of ulterior motives and exaggerated intentions (Silver, Newman, and Small 2021). We found that, for people with lower (vs. higher) trust in business, the impact appeal elicited stronger perceptions of ulterior motives (interaction b

= -.233, $p < .001$) and stronger perceptions of exaggeration (interaction b = -.429, $p < .001$). Furthermore, in univariate regressions, ulterior motives (b = -.184, $p < .001$) and exaggeration (b = -.290, $p < .001$) both predicted weaker survey taking intentions. Importantly, however, when including both of these variables in a multivariate regression predicting survey taking intentions, only exaggeration (b = -.296, $p < .001$) remained statistically significant (ulterior motives b = .011, $p = .733$). These results suggest that the impact appeal may have had more negative effects on survey taking intentions among people with lower trust in business because these people perceived that the company using the impact appeal was intentionally exaggerating the impact customers could make, not because they perceived the company was using the impact appeal for other reasons (e.g., to look good).

As a robustness check, we confirmed that our moderated mediation models hold after controlling for the alternative mediators: perceived impact, perceived threat to freedom, feedback appreciation, personal benefits, and personal costs. We also confirmed that the primary moderation results (Hypotheses 1a and 2) were robust to the inclusion of age and gender controls (and if needed in the case of H1a and H2, their interactions with the indicator of the Impact (vs. Control) condition). See the results of the robustness checks in Web Appendices I-K.

Discussion

The results of Study 2 support all three hypotheses. We conceptually replicate the results of the field experiment, showing that the effect of the impact appeal (compared to the Control condition) on customers' intentions to provide feedback was more negative among customers with lower trust in business (Hypothesis 2). In support of the proposed mechanism underlying this effect, the impact appeal was perceived as more inauthentic than the control subject line among customers with lower trust in business (Hypothesis 1a), and perceived inauthenticity negatively predicted feedback provision intentions (Hypothesis 1b). Finally, our moderated mediation analysis provides evidence consistent with our full model that the impact appeal decreased survey taking intentions among customers with low trust in business because these customers perceived the appeal as inauthentic. However, customers with high trust in business did not perceive the impact appeal as particularly inauthentic, and as a result, the appeal did not reduce their survey taking intentions.

In an additional pre-registered study with a similar experimental design (N = 1,505; see Web Appendix L), we replicated these results, and observed that general interpersonal trust (Levine et al. 2018) did not statistically significantly moderate the effects of the impact appeal on survey completion intentions. This suggests that our findings are not simply driven by a general propensity to trust.

To provide a deeper understanding of our key mechanism, we examined two sources of inauthenticity in Study 2: ulterior motives and exaggeration. Perceptions that the firm was intentionally exaggerating customers' potential impact, not that the firm had ulterior motives, helped explain the impact appeal's negative conditional effects on survey taking intentions. Given that this conclusion is based on correlational results, we conducted a supplemental pre-registered study (N = 1,999; see Web Appendix M) to causally test the role of perceived exaggeration by manipulating the plausibility of the stated impact in the appeal. Specifically, in addition to manipulating the type of feedback request, we varied whether participants learned that the feedback they had provided to a company in the past either had or had not made an impact. When people learned that their past feedback had been neglected (and thus had reason to suspect the stated impact in the appeal was exaggerated), impact appeals again increased perceptions of inauthenticity and failed to improve feedback provision intentions; importantly, when people learned that their past feedback had made an impact, the effect of impact appeals on perceived inauthenticity was substantially attenuated, and impact appeals actually increased feedback provision intentions. Together, perceptions that the firm is exaggerating intentions seem to be a strong determinant of the inauthenticity and inefficacy of impact appeals.

STUDY 3: MANIPULATING TRUST IN BUSINESS

The primary purpose of Study 3 is to use an exogenous manipulation of trust in business to show that it *causally* moderates the effect of impact appeals on survey taking intentions. The study has two additional goals. First, to demonstrate generalizability across various impact appeals, the study incorporated a wider range of stimuli; participants responded to three different impact appeals and three different control subject lines. Second, to demonstrate generalizability across contexts, the feedback request was for a company's product, rather than about a customer service experience.

Methods

Following our pre-registration, we recruited 1,007 participants on Amazon's Mechanical Turk who lived in the USA (mean age = 39.6, SD age =12.2, 58.2% female).

Participants were randomly assigned to either the "Reduced Trust" or "Status Quo" condition. In the Reduced Trust condition, participants were asked to read about declining trust in business from reports by leading market research firms (Edelman and Ipsos). Participants were then asked to answer one comprehension check question and to briefly describe marketing practices that had decreased their own trust in companies in general. In the Status Quo condition, participants were asked to read about the 2022 Toyota Camry, answer one comprehension check question, and briefly describe some car features that they frequently use (or if they do not drive, car features that they believe others frequently use). The Reduced Trust (vs. Status Quo) manipulation was adapted from previous work (Darke and Argo 2005; Kirmani and Zhu 2007).

In the next section of the survey, participants were asked to imagine themselves as a customer of a large technology company referred to as "Company X." Participants were then asked to imagine that on that day, Company X emailed them requesting feedback about its software, and Company X was considering which subject line to use for this email. Participants reacted to three pairs of subject lines, with each pair containing one control subject line and one impact appeal subject line (see Table 2).

Table 3-2. Subject Lines from Study 3

Pair	Control Subject Line	Impact Appeal Subject Line
1	Company X customer feedback survey invitation	Your voice is important: Have a say in Company X's direction
2	Company X requests your feedback	Your opinion matters: Shape the future of Company X
3	Company X customer feedback request	Your expert advice will impact Company X

Participants first answered three questions measuring *survey taking intentions* (one for each pair): "Would you be more likely to take the feedback survey if Company X actually sent you this email with subject line (a) or subject line (b)?" (1 – Much more likely with (a); 7 – Much more likely with (b)). Then participants responded to three questions measuring *perceived inauthenticity* (one for each pair): "If Company X actually sent you this email… which of these two subject lines would feel more inauthentic?" (1 – (a) would feel much more inauthentic; 7 – (b) would feel much more inauthentic).

The order of the pairs was randomized, and each participant was randomized to see the impact appeals either consistently on the left side of the scales (i.e., designated as the subject line (a)) or consistently on the right side of the scales (i.e., designated as the subject line (b)). Each pair's control subject line was always on the opposite side of the scale from its impact appeal subject line. The responses were re-coded such that higher values correspond with the impact appeal subject lines and lower values correspond with the control subject lines. That is, for the measure of survey taking intentions, higher values indicate that participants more strongly preferred to take the survey with an impact appeal over its corresponding control subject line; for the measure of perceived inauthenticity, higher values indicate that participants more strongly believed an impact appeal was more inauthentic than its corresponding control subject line. Participants then completed a two-item trust manipulation check (Rajavi, Kushwaha, and Steenkamp 2019). The two items were highly correlated (Pearson $r = .83$) so we averaged them to create a composite measure of *trust in Company X*. Participants then reported their gender and age.

Results

We first confirmed that the Reduced Trust (vs. Status Quo) manipulation reduced participants' trust in Company X (b = -.565, $p < .001$).

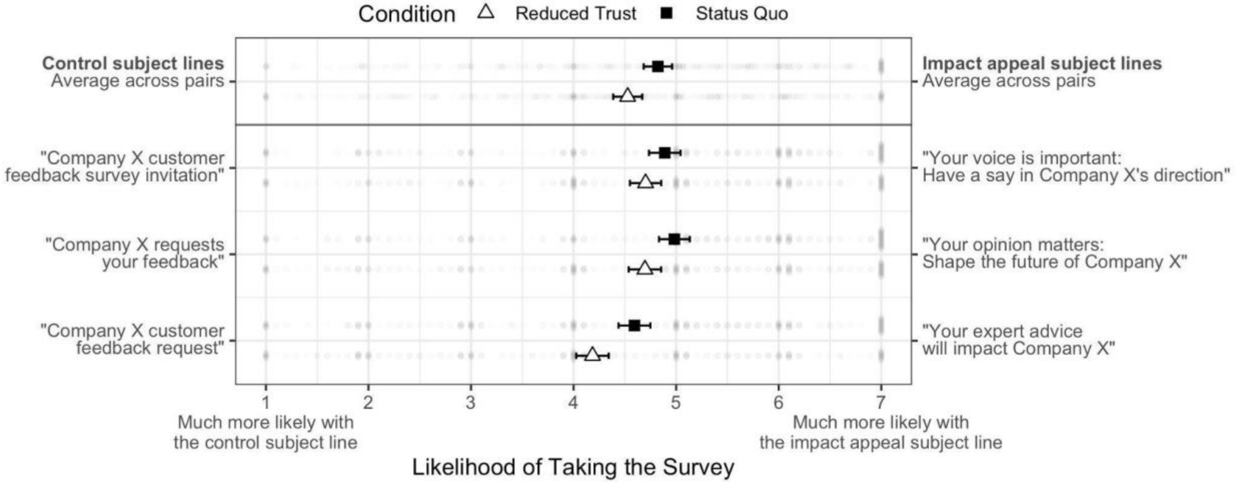
To test our three hypotheses, we analyzed data at the subject line pair level with three observations per participant, using OLS regressions with subject line pair fixed effects and standard errors clustered by participant. We first tested Hypothesis 2 by examining whether prompting (vs. not prompting) people to have lower trust in business decreased their preference for taking a feedback survey with an impact appeal subject line (relative to a survey with a control subject line). The regression predicts survey taking intentions as the outcome, and the right-hand side includes a dummy variable indicating whether participants were in the Reduced Trust (vs. Status Quo) condition. Consistent with Hypothesis 2, being randomly assigned to the Reduced Trust (vs. Status Quo) condition decreased the extent to which people preferred taking a survey with an impact appeal subject line over a control subject line (b = -.294, *p* = .004; Figure 3-5, Panel A).

We next tested Hypothesis 1a that prompting (vs. not prompting) people to have lower trust in business increased the extent to which an impact appeal subject line was viewed as more inauthentic (relative to a control subject line). The regression predicts perceived inauthenticity as the outcome, and the right-hand side includes a dummy variable indicating whether participants were in the Reduced Trust (vs. Status Quo) condition. Consistent with Hypothesis 1a, being randomly assigned to the Reduced Trust (vs. Status Quo) condition increased the extent to which the impact appeal subject lines were viewed as more inauthentic relative to control subject lines (b = .391, *p* < .001; Figure 3-5, Panel B).

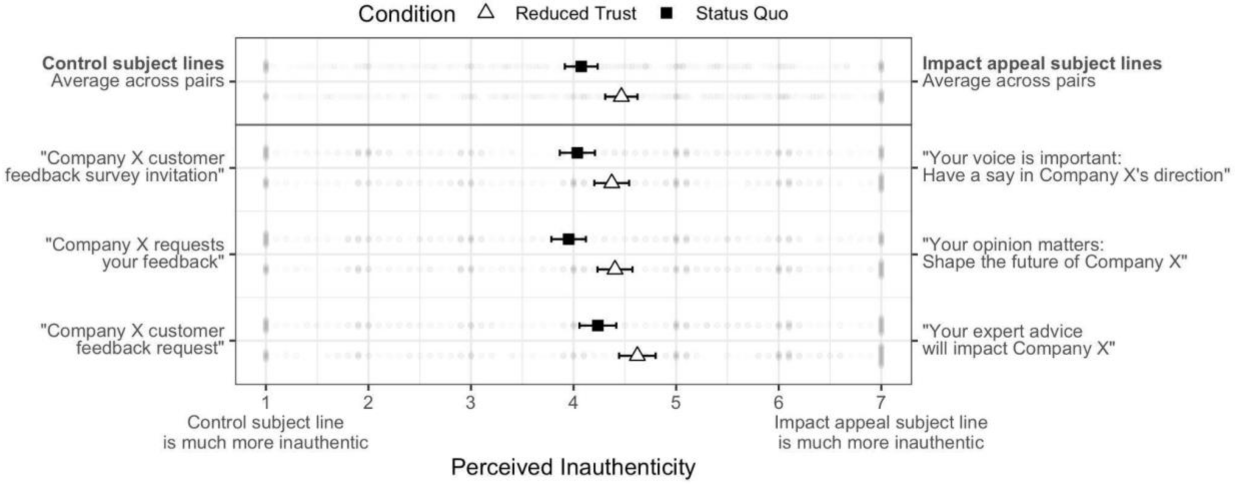
Figure 3-5. The Effects of Trust in Business on Survey Taking Intentions and Inauthenticity in Study 3.

This figure shows participants' responses for survey taking intentions (Panel A) and perceived inauthenticity (Panel B) by trust condition. The square and triangular markers reflect means, error bars reflect 95% confidence intervals, and translucent points represent each participant's response. The top row of each graph displays each participant's average response across the three pairs of subject lines. In the bottom three rows of each graph, the left side displays the exact wording of the control subject lines, and the right side displays the exact wording of the impact appeal subject lines.

Panel A. Survey Taking Intentions



Panel B. Perceived Inauthenticity



We further tested whether the effects of the trust manipulation on survey taking intentions were explained by perceived inauthenticity. We first confirmed that, in support of Hypothesis 1b, the more impact appeals were perceived as inauthentic relative to control subject lines, the less likely people preferred to take surveys with impact appeals over control subject lines (b = -.374, $p < .001$). We then used mediation analysis with 5,000 bootstrapped samples and found that the Reduced Trust (vs. Status Quo) manipulation had a negative indirect effect on survey taking intentions via perceived inauthenticity (indirect effect = -.161, 95% CI = [-.257, -.068]). These results suggest that reducing trust in business increases the extent to which people perceive impact appeal subject lines as more inauthentic than control subject lines, which predicts weaker intentions to take a survey with an impact appeal (vs. a control) subject line.

Discussion

Study 3 provides evidence that trust in business causally moderates the effects of impact appeals. The results support our full theoretical account that impact appeals have more negative effects on feedback provision when customers have lower trust in business because reducing trust causes these customers to perceive impact appeals as more inauthentic. The results also highlight the generalizability of our theorizing; the hypothesized effects were observed across a variety of impact appeals and in response to feedback requests about products (rather than only customer service).

GENERAL DISCUSSION

Customer feedback is a key input into marketing research and strategy. Marketers often try to solicit this feedback by attempting to persuade customers that their actions can impact the organization—a tactic we term “impact appeals.” Previous research and common wisdom among marketers suggest that impact appeals should be effective in increasing customers’ compliance with feedback requests. We tested this hypothesis using a multinational field experiment with a Fortune 500 company, involving over 400,000 customers in seven countries (Study 1). In contrast to our initial prediction and the predictions of marketing experts, impact appeals on average *reduced* compliance with feedback requests. Importantly, the effects varied across countries. We theorized and showed evidence that the effect of impact appeals on feedback provision depends on people's trust in business (Hypothesis 2). Specifically, in the field experiment, we found that impact appeals had more negative effects in countries with lower trust in business (e.g., Japan) than in countries with higher trust in business (e.g., China). In two lab studies, we showed that trust in business (measured in Study 2 and manipulated in Study 3) moderated the effect of impact appeals on compliance with feedback requests. We moreover found support for our proposed underlying mechanism: When people have lower trust in business, they view impact appeals as more inauthentic (Hypothesis 1a), which reduces their willingness to offer feedback in response to impact appeals (Hypothesis 1b).

Our research contributes to the literatures on customer feedback, customer empowerment, and social influence. While previous research has largely focused on the consequences of customer feedback provision (e.g., Dholakia 2010; Lemon and Verhoef 2016), we provide insights into the antecedents of feedback provision. To do this, we conducted high-powered field and lab tests of an influence tactic that is commonly used to solicit feedback but unexplored in previous research. Although existing research suggests that people are motivated by self-efficacy and that people value having impact (e.g., Cryder, Loewenstein, and Scheines 2013; Fuchs and Schreier 2011), we find that prospectively framing an action as empowering can reduce motivation to take that action. To understand this backfiring effect, we identify perceived inauthenticity —in our case, driven by the inference that the firm is exaggerating impact—as a mechanism that explains how influence tactics reduce compliance, and we provide convergent evidence on how trust in business shapes customers' responses to influence tactics.

Managerial Implications

Our work has several managerial implications. The results first and foremost caution against using impact appeals to motivate customer feedback provision without accounting for how the appeals will be perceived by the target audience. In our surveys, marketing experts predicted that "Your voice is important: Shape the [Company] customer experience" would be the most effective subject line in increasing compliance with the feedback requests. If our field partner were to implement this subject line (rather than the control subject line) across the seven countries in our study, a back-of-the-envelope calculation suggests that this would result in approximately 35,000 fewer surveys completed in a year (see Web Appendix N for details). This underscores the importance of testing marketing strategies in the field, even seemingly attractive ones, before scaling them.

Our work further highlights that firms should attend to heterogeneous responses to their marketing strategies. The same impact appeal that effectively mobilizes customers with high trust in business disengages customers with low trust in business. Thus, a targeted approach to soliciting customer feedback—sending impact appeals to customers with high trust in business and straightforward requests to customers with low trust in business—may be the most effective strategy. Identifying moderators like trust in business further enables out-of-sample predictions that can inform broader strategies for firms. In our field partner's case, for the countries in our experiment (accounting for approximately 65% of the total customer base), the firm can use the subject line that performed best in our experiment within each country. For the countries not included in the field experiment (e.g., India, South Korea), the firm can *predict* which subject line would perform best (using the estimated moderation model from the field experiment) and implement that subject line within each out-of-sample country. This targeted approach would, for example, prescribe using an impact appeal subject line in India (with a high level of trust in business, similar to China) and the neutral control subject line in South Korea (with a low level of trust in business, similar to Japan). If our field partner were to use such a targeted approach based on trust in business across its entire customer base, a back-of-the-envelope calculation (see details in Web Appendix N) suggests it would collect approximately 40,000 additional completed surveys over a year compared to only using the control subject line (a 5% relative increase) or 65,000 more surveys compared to only using impact appeals (an 8% relative increase). This exercise shows the value of identifying key moderators that explain heterogeneous treatment effects, which can inform targeted strategies that extend beyond the sample used in a specific field experiment. We encourage future research to more formally assess the value of our recommended approach.

Finally, our theoretical account directly informs ways to redesign impact appeals to effectively mobilize a broader segment of customers. Our research suggests that impact appeals seem inauthentic and backfire if customers feel like the organization is exaggerating potential customer impact. This mechanism suggests that impact appeals could feel motivating, as we originally theorized—if the stated impact seems plausible to customers. As one example, in a supplemental pre-registered study (N = 1,417; see Web Appendix O), we show that impact appeals have positive effects on survey taking intentions if they are complemented with "process transparency information" that describes *how* the firm actually uses customer feedback to make impact (Buell, Porter, and Norton 2021). Future research is needed to test whether alternative implementations of impact appeals that guard against perceptions of inauthenticity—such as process transparency information—could lead to greater feedback provision across a broad audience, even among customers with low general trust in business.

Limitations and Future Research

The current research has several limitations that open up avenues for future research. First, the analysis of trust in business as a moderator in the field was conducted post-hoc. Future research should deductively field test whether the effects of impact appeals and other influence tactics depend on trust in business. Another limitation of this research is that the theoretical account focuses on one reason why impact appeals can backfire; however, there could be other mechanisms involved. For instance, some customers may simply not want to make an impact and instead prefer to leave consequential decisions to experts (Kassirer, Levine, and Gaertig 2020). Future research should examine other mechanisms underlying the effects of impact appeals.

Conclusion

In this research, we find that impact appeals—that is, appeals that attempt to motivate customers by telling them that their actions can impact the organization—can surprisingly decrease customers' willingness to comply with feedback requests. We show that impact appeals can backfire because they are perceived as inauthentic, particularly when customers have lower trust in business. While the belief that one's actions are consequential may underlie customer motivation and engagement in many domains, effectively instilling that belief in customers may be more difficult for firms than it seems.

CHAPTER 3

APPENDIX

Web Appendix A: Surveys on Expert Predictions and Use of Feedback Request Strategies

Survey 1. Customer Experience Professionals' Predictions and Use of Strategies

Methods

We collaborated with the Customer Experience Professional Association (CXPA) to recruit their members and social media followers for a survey via weekly email newsletters and social media posts. As pre-registered, we only included participants in our analysis who (1) completed the survey within four weeks of its initial launch, (2) reported having experience working in customer experience or marketing, and (3) reported having no prior knowledge about the results of our field experiment. Of the 85 people who responded to the survey, 76 met these inclusion criteria ("full sample"). Of these participants, 42 participants worked in the USA ("USA-based sample").

The survey explained:

> *A large US-based technology company (let's call it "Company X") regularly invites customers to take a feedback survey after customers receive help from customer service. Company X conducted an A/B test to examine how email subject lines influence whether customers are willing to take their survey. Specifically, Company X split their customers into equal groups and emailed each group a feedback survey invitation with a different subject line. For each subject line, Company X measured the "survey response rate" (i.e., the share of customers who started the survey).*
>
> *As indicated by the slider below, 25.8% of its US-based customers who received an email with the subject line "Company X customer experience survey invitation" started the survey.*

Note that 25.8% is the actual response rate from the control condition of the field experiment (Study 1) for customers in the USA. Participants then predicted the response rates for the four other subject lines among USA-based customers in our field experiment (using sliders that ranged from 0% to 100%, anchored at 0%):

> [Time:] *Company X customer experience survey invitation (only takes 2 minutes)*
> [Impact (Voice):] *Your voice is important: Shape the Company X customer experience*
> [Impact (Help):] *Your help is needed: Shape the Company X customer experience*
> [Impact (Expert):] *Your expert advice is appreciated: Shape the Company X customer experience*

Note that the words in brackets were not shown to survey respondents and are displayed here as shorthand labels of these subject lines. The order of the three impact appeals was randomized.

Participants were then asked whether they had experience working in marketing or customer experience, whether they were currently working in the USA, and whether they had prior knowledge about the results of our field experiment.

Next, participants were asked, "Based on your observations, which of the following approaches does your company most frequently prefer to use to request feedback from customers?". They had four response options:

1. [Impact appeals:] We try to persuade customers that their feedback will have an impact (for example, "Shape our company's customer service experience"). We do not offer rewards for providing feedback.
2. [Neutral control:] We send straightforward requests (for example, "Customer experience survey invitation"). We do not offer rewards for providing feedback.
3. [Reward:] We offer rewards for completing surveys (for example, "Take this survey to get a gift card").
4. I do not know / Not applicable.

Again, the words in brackets were not shown to survey respondents and are displayed here as shorthand labels of the different types of feedback request strategies. The order of the first three options was randomized.

Finally, participants who thought their company preferred using impact appeals were asked to briefly describe why they think their company prefers this approach.

Results

We pre-registered to calculate each participant's *predicted response rate* for impact appeals (by averaging across their predictions for the three impact appeal subject lines) and to then conduct a t-test comparing the predicted response rates for impact appeals to 25.8% (i.e., the response rate provided to participants for the control subject line).

Our primary pre-registered analysis only focused on the USA-based sample (N=42). The average predicted response rate for impact appeals (29.5%) was marginally significantly greater than 25.8% ($p = .051$). As a secondary analysis, we separately compared participants' predictions for each of the four subject lines with the 25.8% response rate for the control subject line. On average, participants predicted that Time (31.4%; $p = .021$), Impact (Voice) (32.0%; $p = .010$), and Impact (Expert) (30.5%; $p = .019$) would significantly increase response rates relative to the control subject line, and that the Impact (Help) subject line (25.9%; $p = .980$) would not significantly affect response rates relative to the control.

As a pre-registered robustness check we conducted the same comparisons with the full sample (N=76) and found that the average predicted response rate for the impact appeals (31.2%) was significantly greater than 25.8% ($p < .001$). In this larger sample, participants on average predicted that all the treatment subject lines would (significantly or marginally significantly) increase response rates relative to the control subject line: Time (31.9%; $p < .001$), Impact (Voice) (32.6%; $p < .001$), Impact (Expert) (31.3%; $p = .002$), and Impact (Help) (29.0%; $p = .063$).

In the USA-based sample, 50% of participants reported that their firm most frequently preferred using impact appeals, 28.6% reported using the neutral control strategy, 16.7% reported using rewards, and 4.8% reported "I don't know / Not applicable." In the full sample, 48.7% of participants reported that their firm most frequently preferred using impact appeals, 32.9% reported using the neutral control strategy, 11.8% reported using rewards, and 6.6% reported "I don't know / Not applicable."

Survey 2. Customers' Exposure to Feedback Request Strategies

Methods

We recruited 201 participants on Amazon's Mechanical Turk for this survey (mean age = 38.3, SD age = 10.1; 41.8% Female). Participants first learned that the study was about their past experiences with customer feedback requests from companies. Participants were then asked: "In the past few years, how often have you seen companies use each of the following email strategies?" (1 – Never to 6 – Very frequently):

1. [Impact appeal:] They try to persuade customers that their feedback will have an impact (for example, "Your voice is important: Shape the [Company] customer experience"). There is no reward for providing feedback.
2. [Neutral control:] They simply send a straightforward request (for example, "[Company] customer experience survey invitation"). There is no reward for providing feedback.
3. [Reward:] They offer a reward for completing the survey (for example, "Take this survey to get a gift card").

Next, they were asked to search their email inbox for "the most recent email from a company that is requesting your customer feedback." They then reported the email's subject line and rated whether it was an impact appeal (similar to the definition provided above) and whether it offered a reward. They next were asked to read the email body and report whether the email body included an impact appeal and whether it offered a reward. They provided their age and gender at the end of the study.

Results

The responses to the first question about how often they had seen different strategies are summarized in Table A3-1 below.

Table A3-1. Perceived Frequencies of Customer Feedback Solicitation Strategies

	Mean (SD)	Median
Impact Appeal	4.54 (1.05)	5
Neutral Control	4.56 (1.16)	5
Reward	3.33 (1.44)	3

Note: The scale ranges from: 1 – Never to 6 – Very Frequently.

Of the 201 participants who completed the survey, 83.6% (168) reported that they were able to find a customer feedback request in their email inbox. Of the 168 who found a customer feedback request, impact appeals were reported in 39.9% (67) of subject lines and in 70.2% (118) of email bodies. Reward offers were reported in 17.9% (30) of subject lines and in 25% (42) of email bodies.

Survey 3. MBA Marketing Association Members' Predictions

Methods

In a pre-registered survey, we collaborated with the marketing association at a large business school and got permission to email all of their members. We pre-registered that we would leave the survey open for 14 days and that we would stop data collection at that point. Of the 68 MBA students who took the survey, 94% had completed course work in marketing and 68% had worked in marketing.

The survey explained:

> *A large US-based technology company ("Company X") seeks to increase the amount of customer feedback. About a day after customers receive help from a service agent, the company emails customers, inviting them to take a voluntary customer experience survey. The company is considering using one of the five following email subject lines for survey invitations sent to their U.S. customers. In this context, which subject line do you think will result in the highest rate of survey completion among the invited customers?*

Participants had 5 options:

> [Control:] *Company X customer experience survey invitation*
> [Time:] *Company X customer experience survey invitation (only takes 2 minutes)*
> [Impact (Voice):] *Your voice is important: Shape the Company X customer experience*
> [Impact (Help):] *Your help is needed: Shape the Company X customer experience*
> [Impact (Expert):] *Your expert advice is appreciated: Shape the Company X customer experience*

Note that the words in brackets were not shown to survey respondents and are displayed here as shorthand labels of these subject lines.

Results

Of the 68 MBAs who completed the survey, 41% (28) predicted that the Impact (Voice) subject line would lead to the highest consumer feedback survey completion rate, 34% (23) predicted Time, 18% (12) predicted Impact (Help), 6% (4) predicted Impact (Expert), and 1% (1) predicted Control.

More formally, we pre-registered that we would estimate a multinomial logit with the 5-option choice as the dependent variable (DV) and only constant terms on the right-hand side. For the DV, the reference group is the Control subject line. Consistent with our initial hypothesis, we found that, compared to the Control subject line, people were either significantly or at least directionally more likely to pick the Impact (Voice) subject line (b = 3.33, p = .002), the Impact (Help) subject line (b = 2.49, p = .017), and the Impact (Expert) subject line (b = 1.39, p = .215).

We confirmed that the patterns of predictions were similar among MBAs who had previously worked in marketing (N = 46). Specifically, among this subsample, 41% (19) predicted Impact (Voice), 37% (17) predicted Time, 15% (7) predicted Impact (Help), 7% (3) predicted Impact (Expert), and 0% (0) predicted Control would outperform the other subject lines. Also, when we added a binary indicator of whether a respondent had previously worked in marketing as an independent variable in the aforementioned multinomial logit, we did not find evidence that it predicted respondents' choices (p-values > .850).

Web Appendix B: Subject Line Translations (Study 1)

Note that all translations included the actual company name (indicated as “[Company]” below) in English. The company confirmed that its consumers across the world typically use the English version of the company name.

Table A3-2. Subject Line Translations

Experimental Condition	Language					
	English	French	German	Japanese	Portuguese (Brazilian)	Chinese (Simplified)
Control	[Company] customer experience survey invitation	Enquête de satisfaction client [Company]	Einladung zur Umfrage zur Zufriedenheit von [Company] Kunden	[Company]お客様満足度アンケートのご案内	Pesquisa de satisfação referente ao atendimento da [Company]	[Company]诚邀您参加客户体验调查
Time	[Company] customer experience survey invitation (only takes 2 minutes)	Enquête de satisfaction client [Company] (ne prendra que 2mn)	Einladung zur Umfrage zur Zufriedenheit von [Company] Kunden (dauert nur 2 Minuten)	[Company]お客様満足度アンケートのご案内（所要時間約2分）	Pesquisa de satisfação referente ao atendimento da [Company] (leva só dois minutos)	[Company]诚邀您参加客户体验调查(仅需2分钟)
Impact (Voice)	Your voice is important: Shape the [Company] customer experience	Votre voix est importante: Façonnez l'expérience client [Company]	Ihre Stimme ist wichtig: Prägen Sie das [Company] Kundenerlebnis	[Company]の顧客体験をより良くするために、お客様の声を大切にいたします	Sua opinião é importante: Molde a experiência do cliente [Company]	您的意见对[Company]很重要：共建客户体验
Impact (Help)	Your help is needed: Shape the [Company] customer experience	Votre aide est requise: Façonnez l'expérience client [Company]	Ihre Hilfe wird benötigt: Prägen Sie das [Company] Kundenerlebnis	[Company]の顧客体験をより良くするために、お客様のご協力が必要です	Precisamos da sua ajuda: Molde a experiência do cliente [Company]	[Company]需要您的帮助：共建客户体验
Impact (Expert)	Your expert advice is appreciated: Shape the [Company] customer experience	Votre avis d'expert est apprécié: Façonnez l'expérience client [Company]	Ihre Expertise wird geschätzt: Prägen Sie das [Company] Kundenerlebnis	[Company]の顧客体験をより良くするために、お客様からの貴重なアドバイスをお願いいたします	Valorizamos sua expertise: Molde a experiência do cliente [Company]	[Company]十分重视您的专业建议：共建客户体验

Web Appendix C: Subject Line Pre-Test (Study 1)

We recruited participants on Prolific to evaluate the subject lines. For non-English subject lines, we recruited bi-lingual participants who were fluent in English and the respective language. Of the 2,134 participants who started the pre-test survey, 1,336 participants passed all three attention checks. 382 participants pre-tested five English subject lines that we did not end up using in the field experiment so we will not report them here. Thus, 954 participants were included in the final analysis (see breakdown by language below).

Table A3-3. Pre-Test Sample Size by Language

Language	Sample Size
English	294
Chinese	147
French	216
German	227
Japanese	12
Portuguese	58
	Total = 954

For English, Chinese, French, and German translations, participants were randomly assigned to read either the control subject line, or one of the three impact appeal subject lines. Because we had smaller samples for Japanese and Portuguese, we used a within-subjects design for those languages, where participants read all four subject lines. Note that the Impact (Expert) subject line in Portuguese that we tested in this pre-test ("Valorizamos seu conselho de especialista: Molde a experiência do cliente [Company]") was slightly different from the one that we used in the field experiment ("Valorizamos sua expertise: Molde a experiência do cliente [Company]") because we learned later that in Portuguese, "expertise" better captured the English meaning of "expert advice" than "conselho de especialista".

All participants responded to our key manipulation check:

> [Impact manipulation check:] To what extent does this subject line indicate that **customers can actually impact the development of customer experience** at [Company]?
> [from 1 (Not at all) to 7 (Very much)]

In Table A3-4, we report the differences between the three impact subject lines and the control subject line in responses to this manipulation check across languages. For English, Chinese, French, and German, we used a simple OLS regression with a dummy variable indicating whether a subject line included an impact appeal. For Portuguese and Japanese, where we had used a within-subjects design, we reshaped the data so that each participant had four observations (one for each subject line they rated). We estimated OLS regressions with participant fixed effects and standard errors clustered by participant.

Table A3-4. Pre-Test Results

	Dependent variable:					
	Impact manipulation check					
	(1)	(2)	(3)	(4)	(5)	(6)
Impact	1.118^{***}	0.712^{**}	0.723^{***}	0.586^{**}	1.139^{**}	1.149^{***}
	(0.185)	(0.268)	(0.197)	(0.181)	(0.348)	(0.211)
Design	Between-subjects				Within-subjects	
Language	English	Chinese	German	French	Japanese	Portuguese
Observations	294	147	227	216	48	232
R^2	0.111	0.047	0.057	0.047	0.412	0.454

$^{*}p<0.05$, $^{**}p<.01$, $^{***}p<0.001$

Note that we also asked about a number of other reactions to our subject lines that are not directly related to our current theory. The data and survey questions are available on this paper's Researchbox page.

Web Appendix D: Balance Checks (Study 1)

Each row in the following table corresponds to one of the baseline sample characteristics. The note below the table explains how each variable is coded. Each cell includes the respective variable's mean in the given condition. For the binary and continuous baseline characteristics in the first five rows, an F-statistic was estimated from an OLS regression predicting the respective baseline characteristic using indicator variables for the four experimental conditions (vs. control condition) on the right-hand side, with heteroskedasticity robust standard errors. For the categorical baseline characteristic measuring consumer type, a single chi-squared test was used to estimate whether the distribution differs across conditions. None of the statistical tests were significant at the 5% level.

Table A3-5. Balance Checks

	Condition					
	Control	Time	Impact (Voice)	Impact (Help)	Impact (Expert)	Test Statistic
in_warranty	0.91	0.91	0.91	0.91	0.91	F = 1.27
premium_account	0.49	0.49	0.49	0.49	0.49	F = 1.86
subsidiary	0.51	0.51	0.51	0.51	0.51	F = 0.95
invitation_count	0.91	0.91	0.91	0.91	0.92	F = 0.63
onsite	0.23	0.22	0.23	0.22	0.23	F = 1.20
consumer_type1	0.48	0.48	0.48	0.48	0.48	χ^2 = 27.88
consumer_type2	0.06	0.06	0.07	0.07	0.07	
consumer_type3	0.17	0.17	0.17	0.17	0.17	
consumer_type4	0.12	0.12	0.12	0.11	0.12	
consumer_type5	0.15	0.15	0.15	0.15	0.15	
consumer_type6	0.02	0.02	0.02	0.02	0.02	

$^{*}p<0.05$, $^{**}p<.01$, $^{***}p<0.001$

Note: The variables are coded as follows, based on customers' status at the time of data extraction:

in_warranty: = 1 if they were in warranty; = 0 otherwise

premium_account: = 1 if they had premium account status; = 0 otherwise
subsidiary: = 1 if they were a customer of a subsidiary company of our field partner; = 0 otherwise
invitation_count: the number of survey invitations they had received in the prior year (continuous)
onsite: =1 if they had received in-person support for the customer service experience they were asked to take a survey about; = 0 otherwise
consumer_type1: = 1 if "CBO Consumer HO"; = 0 otherwise
consumer_type2: = 1 if "CBO G500 [Global Fortune 500]"; = 0 otherwise
consumer_type3: = 1 if "CBO LargeBusiness"; = 0 otherwise
consumer_type4: = 1 if "CBO MediumBusiness"; = 0 otherwise
consumer_type5: = 1 if "CBO SmallBusiness"; = 0 otherwise
consumer_type6: = 1 if "CBO NonSegCountries", "Inter Company", "CBO CommChannels", or NA; = 0 otherwise

Web Appendix E: Primary Regression Tables (Study 1)

Table A3-6 Column (1) reports results from our primary regression specification: An OLS regression predicting survey starting with two indicator variables (an indicator variable for the Time condition and an indicator variable for the three Impact conditions), and language and week fixed effects (FEs). The regression reported in Table A3-6 Column (2) adds the baseline characteristics described in Web Appendix D. Both regressions use heteroskedasticity robust standard errors. Table A3-6 Columns (3)-(6) and Table A3-7 apply the same regression specifications to other outcome variables.

Table A3-6. Average Treatment Effects on Survey Starting, Survey Completion, and Unsubscribing

	Dependent variable:					
	Survey starting		Survey completion		Unsubscribing	
	(1)	(2)	(3)	(4)	(5)	(6)
Time	0.003	0.003	0.003	0.003	-0.0001	-0.0001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.0002)	(0.0002)
Impact	-0.004^{*}	-0.004^{*}	-0.009^{***}	-0.009^{***}	0.0004^{*}	0.0004^{*}
	(0.002)	(0.002)	(0.002)	(0.002)	(0.0002)	(0.0002)
Controls	No	Yes	No	Yes	No	Yes
Language FEs	Yes	Yes	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	430,666	430,666	430,666	430,666	430,666	430,666
R^2	0.031	0.046	0.041	0.054	0.001	0.001

$^{*}p<0.05$, $^{**}p<.01$, $^{***}p<0.001$

Table A3-7. Average Treatment Effects on Satisfaction Score Provided by Customers in the Feedback Survey

	Dependent variable:	
	Satisfaction	
	(1)	(2)
Time	0.039	0.030
	(0.028)	(0.027)
Impact	-0.019	-0.025
	(0.023)	(0.023)
Controls	No	Yes
Language FEs	Yes	Yes
Week FEs	Yes	Yes
Observations	108,671	108,671
R^2	0.012	0.074

$^{*}p<0.05$, $^{**}p<.01$, $^{***}p<0.001$

Web Appendix F: Secondary Models and Robustness Checks (Study 1)

1. Separate Impact Appeals

This table includes models that are identical to the primary models (reported in Table A3-6) except that these models include a separate dummy variable for each of the three impact appeal treatment conditions.

Table A3-8. Average Treatment Effects on Survey Starting, Survey Completion, and Unsubscribing (Indicators for Each Condition)

	Dependent variable:					
	Survey starting		Survey completion		Unsubscribing	
	(1)	(2)	(3)	(4)	(5)	(6)
Time	0.003	0.003	0.003	0.003	-0.0001	-0.0001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.0002)	(0.0002)
Impact (Voice)	-0.010***	-0.010***	-0.013***	-0.013***	0.00001	0.00001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.0002)	(0.0002)
Impact (Help)	0.002	0.002	-0.003	-0.003	0.001***	0.001***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.0002)	(0.0002)
Impact (Expert)	-0.005*	-0.005*	-0.010***	-0.010***	0.0002	0.0002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.0002)	(0.0002)
Controls	No	Yes	No	Yes	No	Yes
Language FEs	Yes	Yes	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	430,666	430,666	430,666	430,666	430,666	430,666
R^2	0.031	0.046	0.041	0.054	0.001	0.001

$^{*}p<0.05$, $^{**}p<.01$, $^{***}p<0.001$

2. Logistic Regressions

This table uses logistic regressions (instead of OLS).

Table A3-9. Estimated Treatment Effects Using Logistic Regressions

	Dependent variable:					
	Survey starting		Survey completion		Unsubscribing	
	(1)	(2)	(3)	(4)	(5)	(6)
Time	0.015	0.015	0.017	0.018	-0.042	-0.042
	(0.011)	(0.011)	(0.012)	(0.012)	(0.123)	(0.123)
Impact	-0.022*	-0.023*	-0.055***	-0.056***	0.215*	0.215*
	(0.009)	(0.009)	(0.010)	(0.010)	(0.097)	(0.097)
Controls	No	Yes	No	Yes	No	Yes
Language FEs	Yes	Yes	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	430,666	430,666	430,666	430,666	430,666	430,666

* p<0.05, **p<.01, ***p<0.001

3. Mixed Effects Regressions

This table shows results from mixed effects regressions that include language and week random effects (REs) with varying intercepts (instead of language and week fixed effects).

Table A3-10. Estimated Treatment Effects Using Mixed Effects Regressions

	Dependent variable:					
	Survey starting		Survey completion		Unsubscribing	
	(1)	(2)	(3)	(4)	(5)	(6)
Time	0.003	0.003	0.003	0.003	-0.0001	-0.0001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.0002)	(0.0002)
Impact	-0.004*	-0.004*	-0.009***	-0.009***	0.0004*	0.0004*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.0002)	(0.0002)
Controls	No	Yes	No	Yes	No	Yes
Language Random Effects	Yes	Yes	Yes	Yes	Yes	Yes
Week Random Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	430,666	430,666	430,666	430,666	430,666	430,666

* p<0.05, **p<.01, ***p<0.001

Web Appendix G: Moderation by Trust in Business (Study 1)

Table A3-11 reports results from OLS regressions that include interactions between country-level trust in business (mean centered) and the two indicators for the experimental conditions (*Time* and *Impact*). Note that (non-interacted) country-level trust in business is not in the regressions because it is absorbed by the country fixed effects. Standard errors are clustered by country. Table A3-11 shows the effects on *survey starting, survey completion,* and *unsubscribing.* Tables A3-12-A3-14 report robustness checks that use the same specification and outcomes, except with a measure of country-level trust in business (i) that was collected by Edelman in 2019 before the pandemic (Table A3-12), (ii) that uses the average value of trust within each country (rather than the share of people who trust business within each country; Table A3-13), and (iii) while also controlling for alternative moderators including uncertainty avoidance, power distance, and individualism (Hofstede, Hofstede, and Minkov 2010; Table A3-14). All moderators are mean-centered. In each table, "Controls" refer to baseline characteristics described in Web Appendix D.

Table A3-11. Moderation by Trust in Business (Results Reported in the Main Text)

	Dependent variable:					
	Survey starting		Survey completion		Unsubscribing	
	(1)	(2)	(3)	(4)	(5)	(6)
Time	0.003	0.003	0.003	0.003	-0.0001	-0.0001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.0001)	(0.0001)
Impact	-0.004	-0.004	-0.009**	-0.009**	0.0004**	0.0004**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.0001)	(0.0001)
Trust in business * Time	0.0004*	0.0004*	0.00003	0.00004	-0.00001	-0.00001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.00001)	(0.00001)
Trust in business * Impact	0.002***	0.002***	0.002***	0.002***	-0.00003**	-0.00003**
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.00001)	(0.00001)
Controls	No	Yes	No	Yes	No	Yes
Country FEs	Yes	Yes	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	430,666	430,666	430,666	430,666	430,666	430,666
R^2	0.033	0.047	0.042	0.056	0.001	0.001

* p<0.05, **p<.01, ***p<0.001

Table A3-12. Moderation by Trust in Business (Measured in 2019)

	Dependent variable:					
	Survey starting		Survey completion		Unsubscribing	
	(1)	(2)	(3)	(4)	(5)	(6)
Time	0.003	0.003	0.003	0.003	-0.0001	-0.0001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.0001)	(0.0001)
Impact	-0.004	-0.004	-0.009*	-0.009*	0.0004**	0.0004**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.0001)	(0.0001)
Trust in business (2019) * Time	0.0002**	0.0003*	0.00004	0.00005	-0.00000	-0.00000
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.00000)	(0.00000)
Trust in business (2019) * Impact	0.001***	0.001***	0.001**	0.001**	-0.00002*	-0.00002*
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.00000)	(0.00000)
Controls	No	Yes	No	Yes	No	Yes
Country FEs	Yes	Yes	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	430,666	430,666	430,666	430,666	430,666	430,666
R^2	0.033	0.047	0.042	0.056	0.001	0.001

*p<0.05, **p<.01, ***p<0.001

Table A3-13. Moderation by Trust in Business (Measured as the Average Rating in a Country)

	Dependent variable:					
	Survey starting		Survey completion		Unsubscribing	
	(1)	(2)	(3)	(4)	(5)	(6)
Time	0.003	0.003	0.003	0.003	-0.0001	-0.0001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.0001)	(0.0001)
Impact	-0.004	-0.004	-0.009**	-0.009**	0.0004**	0.0004**
	(0.003)	(0.003)	(0.002)	(0.002)	(0.0001)	(0.0001)
Trust in business (average) * Time	0.010*	0.010	0.002	0.002	-0.0002	-0.0002
	(0.004)	(0.004)	(0.004)	(0.004)	(0.0002)	(0.0002)
Trust in business (average) * Impact	0.051***	0.051***	0.040***	0.040***	-0.001**	-0.001**
	(0.008)	(0.008)	(0.007)	(0.007)	(0.0002)	(0.0001)
Controls	No	Yes	No	Yes	No	Yes
Country FEs	Yes	Yes	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	430,666	430,666	430,666	430,666	430,666	430,666
R^2	0.033	0.047	0.042	0.056	0.001	0.001

* $p<0.05$, **$p<.01$, ***$p<0.001$

Table A3-14. Moderation by Trust in Business (Controlling for Alternative Moderators)

	Dependent variable:					
	Survey starting					
	(1)	(2)	(3)	(4)	(5)	(6)
Time	0.003	0.003	0.003	0.003	0.003	0.003
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Impact	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Trust in business * Time	0.001^{**}	0.001^{**}	-0.0002	-0.0003	-0.0001	-0.0001
	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Trust in business * Impact	0.002^{**}	0.002^{**}	0.003^{***}	0.003^{***}	0.003^{***}	0.003^{***}
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0002)	(0.0002)
Uncertainty avoidance * Time	0.0003^{**}	0.0003^{**}				
	(0.0001)	(0.0001)				
Uncertainty avoidance * Impact	-0.0003	-0.0003				
	(0.0002)	(0.0002)				
Power distance * Time			0.0003	0.0004		
			(0.0002)	(0.0002)		
Power distance * Impact			-0.0004	-0.0004		
			(0.0002)	(0.0002)		
Individualism * Time					-0.0002	-0.0002^{*}
					(0.0001)	(0.0001)
Individualism * Impact					0.0001	0.0001
					(0.0001)	(0.0001)
Controls	No	Yes	No	Yes	No	Yes
Country FEs	Yes	Yes	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	430,666	430,666	430,666	430,666	430,666	430,666
R^2	0.033	0.047	0.033	0.047	0.033	0.047

$^{*}p<0.05$, $^{**}p<.01$, $^{***}p<0.001$

While there seems to be a statistically significant interaction between trust in business and the Time condition in some of the models in Tables A3-11-A3-14, we believe this moderating relationship is not as robust as the interaction between trust in businesses and the Impact conditions for two reasons:

1. **Magnitude of the moderating relationship**. Compared to the Impact conditions, the treatment effect of the Time condition on survey starting is moderated by trust in business to a much weaker extent. For example, in our primary model reported in Table A3-11, Model (1), the coefficient on *Trust in business * Time* (i.e., b = .0004) is only one fifth of the coefficient on *Trust in business * Impact* (i.e., b = .002). Further, when estimating the seven-point correlation between the regression-estimated effects of a subject line (relative to the Control condition) and trust in business across countries, we see that the correlation is quite weak for the Time condition (Pearson *r* = .10), but the correlation is strong for the Impact conditions (Pearson *r* = .73).
2. **Instability across secondary outcome variables and alternative specifications**. In Tables A3-11-A3-13, the interaction between trust in business and the Time condition is never a statically significant predictor of survey completion or unsubscribing. Also of note, in Table A3-13, Model (2), the interaction between trust in business and the Time condition does not statistically significantly predict survey starting. In contrast, the interaction between trust in business and the Impact conditions is statistically significant across all outcome variables and model specifications.

Together, unlike the Impact conditions, the Time condition does not seem to meaningfully and reliably interact with trust in business in predicting the various outcome measures.

Web Appendix H: Measurement of Alternative Mechanisms (Study 2)

Table A3-15. Measurement of Alternative Mechanisms

Scale label	Scale items
Perceived impact (Spreitzer 1995; Fuchs, Prandelli, and Schreier 2010)	1. I have significant influence over [Company]'s customer service experience. 2. I have a large impact on how [Company] provides customer service.
Perceived threat to freedom (Irmak, Murdock, and Kanuri 2020)	1. [Company] is using this subject line in an attempt to control my behavior as a customer. 2. [Company] is using this subject line to try to take away my control over my decisions as a customer.
Feedback appreciation	1. [Company] is using this subject line because it really values each customer's opinion. 2. [Company] is using this subject line because it really cares about each customer's feedback.
Personal benefits	1. To what extent do you think you would personally benefit from taking [Company]’s feedback survey?
Personal costs	1. How much time do you think it would take you to complete this feedback survey?

Web Appendix I: Trust in Business (Study 2)

To measure trust in business we used the following prompt (adapted from Levine et al. 2018).

> The following statements are about your beliefs regarding **for-profit companies** in general.
>
> Please rate the extent to which you agree or disagree with each statement. [1 – Strongly disagree; 6 – Strongly agree]
>
> 1. If a company provided a customer satisfaction guarantee, I believe that the company would honor that guarantee.
> 2. If I purchased an item from a company and it did not work as advertised, I believe the company would offer me a full refund for the item.
> 3. I believe that companies do not intentionally misrepresent their products and services to customers.
> 4. I believe that companies tell customers the truth.

Models 1 and 4 in Table A3-16 report OLS regressions that predict survey taking intentions (Model 1) and perceived inauthenticity (Model 4) as a function of a dummy variable indicating whether participants were in the Impact (vs. Control) condition, (mean-centered) trust in business, and the interaction between these variables. We report these results in the manuscript. Models 2 and 5 also include a covariate for prior feedback provision; however, we interpret these results with caution because this covariate could be causally affected by the moderator (see footnote 10 in the manuscript for more information). As a robustness check, we also controlled for age (mean-centered) and gender (binary) and their interactions with the Impact (vs. Control) condition indicator (i.e., interaction controls) in Models 3 and 6.

Tables A3-17-A3-19 report OLS regressions that predict secondary mechanisms as a function of a dummy variable indicating whether participants were in the Impact (vs. Control) condition, trust in business (mean-centered), and the interaction between these variables. Each outcome is estimated once without controls and once controlling for age (mean-centered) and gender (binary) and their interactions with the Impact (vs. Control) condition indicator (i.e., interaction controls).

Table A3-16. Trust in Business Moderation

	Dependent variable:					
	Survey taking intentions			Perceived Inauthenticity		
	(1)	(2)	(3)	(4)	(5)	(6)
Impact	-0.033	0.049	-0.051	0.475^{***}	0.470^{***}	0.479^{***}
	(0.061)	(0.047)	(0.059)	(0.050)	(0.050)	(0.050)
Trust in business	0.449^{***}	0.217^{***}	0.481^{***}	-0.104^{*}	-0.090^{*}	-0.163^{**}
	(0.051)	(0.040)	(0.065)	(0.042)	(0.042)	(0.054)
Impact * Trust in business	0.180^{*}	0.109^{*}	0.151^{*}	-0.257^{***}	-0.253^{***}	-0.237^{***}
	(0.070)	(0.054)	(0.069)	(0.057)	(0.057)	(0.057)
Prior Feedback Covariate	No	Yes	No	No	Yes	No
Controls & Interaction Controls	No	No	Yes	No	No	Yes
Observations	1,998	1,998	1,998	1,998	1,998	1,998
R^2	0.111	0.475	0.161	0.083	0.085	0.095

* p<0.05, **p<.01, ***p<0.001

Table A3-17. Trust in Business Moderation of Exaggeration and Ulterior Motives

	Dependent variable:			
	Exaggeration		Ulterior motives	
	(1)	(2)	(3)	(4)
Impact	1.020***	1.024***	0.828***	0.832***
	(0.051)	(0.051)	(0.050)	(0.050)
Trust in business	-0.062	-0.094	-0.071	-0.130*
	(0.043)	(0.055)	(0.042)	(0.055)
Impact * Trust in business	-0.429***	-0.413***	-0.233***	-0.209***
	(0.058)	(0.059)	(0.058)	(0.058)
Controls and Interaction Controls	No	Yes	No	Yes
Observations	1,998	1,998	1,998	1,998
R^2	0.217	0.225	0.142	0.158

* $p<0.05$, $^{**}p<.01$, $^{***}p<0.001$

Table A3-18. Trust in Business Moderation of Alternative Mechanisms (Part 1)

	Dependent variable:			
	Perceived impact		Feedback appreciation	
	(1)	(2)	(3)	(4)
Impact	0.016	0.021	-0.142***	-0.145***
	(0.049)	(0.049)	(0.039)	(0.039)
Trust in business	0.550***	0.559***	0.501***	0.519***
	(0.041)	(0.053)	(0.033)	(0.043)
Impact * Trust in business	0.125*	0.123*	0.173***	0.163***
	(0.056)	(0.056)	(0.045)	(0.045)
Controls and Interaction Controls	No	Yes	No	Yes
Observations	1,998	1,998	1,998	1,998
R^2	0.197	0.203	0.265	0.276

* p<0.05, **p<.01, ***p<0.001

Table A3-19. Trust in Business Moderation of Alternative Mechanisms (Part 2)

	Dependent variable:					
	Perceived threat to freedom		Personal costs		Personal benefits	
	(1)	(2)	(3)	(4)	(5)	(6)
Impact	0.485***	0.491***	0.062	0.072	0.071	0.070
	(0.047)	(0.047)	(0.048)	(0.048)	(0.058)	(0.058)
Trust in business	-0.035	-0.041	-0.033	-0.053	0.516***	0.461***
	(0.039)	(0.051)	(0.040)	(0.052)	(0.048)	(0.063)
Impact * Trust in business	-0.248***	-0.238***	-0.022	-0.005	0.078	0.089
	(0.054)	(0.054)	(0.055)	(0.055)	(0.066)	(0.067)
Controls and Interaction Controls	No	Yes	No	Yes	No	Yes
Observations	1,998	1,998	1,998	1,998	1,998	1,998
R^2	0.077	0.089	0.002	0.020	0.127	0.128

* $p<0.05$, **$p<.01$, ***$p<0.001$

Web Appendix J: Trust in the Specific Company (Study 2)

To measure trust in the specific company we used the following prompt (adapted from Levine et al. 2018).

> The following statements are about your beliefs regarding **[Company] based on your real life experiences as their customer in the past.**
>
> Please rate the extent to which you agree or disagree with each statement. [1 – Strongly disagree; 6 – Strongly agree]
>
> 1. If [Company] provided a customer satisfaction guarantee, I believe that [Company] would honor that guarantee.
> 2. If I purchased an item from [Company] and it did not work as advertised, I believe [Company] would offer me a full refund for the item.
> 3. I believe that [Company] does not intentionally misrepresent their products and services to customers.
> 4. I believe that [Company] tells customers the truth.

Models 1 and 4 in Table A3-20 report OLS regressions that predict survey taking intentions (Model 1) and perceived inauthenticity (Model 4) as a function of a dummy variable indicating whether participants were in the Impact (vs. Control) condition, trust in the specific company (mean-centered), and the interaction between these variables. Models 2 and 5 also include a covariate for prior feedback provision; however, we interpret these results with caution because this covariate could be causally affected by the moderator (see footnote 10 in the manuscript for more information). As a robustness check, we also controlled for age (mean-centered) and gender (binary) and their interactions with the Impact (vs. Control) condition indicator (i.e., interaction controls) in Models 3 and 6.

Table A3-20. Trust in Specific Company Moderation

	Dependent variable:					
	Survey taking intentions			Perceived inauthenticity		
	(1)	(2)	(3)	(4)	(5)	(6)
Impact	0.008	0.071	-0.016	0.450***	0.449***	0.454***
	(0.060)	(0.046)	(0.058)	(0.048)	(0.048)	(0.048)
Trust in specific company	0.589***	0.351***	0.591***	-0.330***	-0.326***	-0.408***
	(0.053)	(0.042)	(0.065)	(0.043)	(0.044)	(0.054)
Impact * Trust in specific company	0.138	0.062	0.125	-0.193**	-0.191**	-0.184**
	(0.073)	(0.056)	(0.071)	(0.059)	(0.059)	(0.058)
Prior Feedback Covariate	No	Yes	No	No	Yes	No
Controls and Interaction Controls	No	No	Yes	No	No	Yes
Observations	1,998	1,998	1,998	1,998	1,998	1,998
R^2	0.145	0.494	0.196	0.140	0.140	0.154

* p<0.05, **p<.01, ***p<0.001

Note, the coefficients on the interaction term in columns (1) and (3) are marginally statistically significant (*p* = .059 and *p* = .077, respectively).

Web Appendix K: Perceived Inauthenticity and Survey Taking Intentions (Study 2)

Model 1 in Table A3-21 reports an OLS regression that predicts survey taking intentions as a function of perceived inauthenticity. We report this result in the manuscript. Model 2 adds a covariate for prior feedback provision. Model 3 adds gender, age, and all alternative mechanisms. Models 1-3 in Table A3-22 respectively report OLS regressions that predict survey taking intentions as a function of exaggeration, ulterior motives, and both. Model 4 in Table A3-22 includes a covariate for prior feedback provision, and Model 5 adds gender, age, and all alternative mechanisms.

Table A3-21. Perceived Inauthenticity and Survey Taking Intentions

	Dependent variable:		
	Survey taking intentions		
	(1)	(2)	(3)
Perceived inauthenticity	-0.381***	-0.298***	-0.181***
	(0.026)	(0.020)	(0.029)
Prior feedback		0.785***	
		(0.019)	
Perceived impact			0.134***
			(0.029)
Feedback appreciation			0.283***
			(0.036)
Perceived threat to freedom			0.040
			(0.030)
Personal costs			-0.144***
			(0.026)
Personal benefits			0.307***
			(0.023)
Age			0.019***
			(0.002)
Female			0.203***
			(0.052)
Observations	1,998	1,998	1,998
R^2	0.094	0.505	0.385

* p<0.05, **p<.01, ***p<0.001

After including all of the controls (those included in Model 3 in Table A3-21), we replicated the moderated mediation analyses reported in the manuscript. Using 5,000 bootstrapped samples, we estimated that the impact appeal had a negative *indirect* effect on survey taking intentions via perceived inauthenticity among people with moderately low trust in business (who on average indicated “disagree” on the trust in business items; indirect effect = -.198, 95% CI = [-.282, -.133]). Yet for people with moderately high trust in business (who on average indicated “agree” on the trust in business items), the indirect effect of the impact appeals through perceived inauthenticity was significantly weaker (indirect effect = -.040, 95% CI = [-.080, -.005]; comparison of indirect effects between the two levels of trust, $p < .001$).

Table A3-22. Exaggeration, Ulterior Motives, and Survey Taking Intentions

	Dependent variable:				
	Survey taking intentions				
	(1)	(2)	(3)	(4)	(5)
Exaggeration	-0.290***		-0.296***	-0.218***	-0.079**[29]
	(0.024)		(0.031)	(0.023)	(0.030)
Ulterior motives		-0.184***	0.011	0.049*	0.057*
		(0.026)	(0.033)	(0.025)	(0.028)
Prior feedback				0.788***	
				(0.020)	
Perceived impact					0.133***
					(0.030)
Feedback appreciation					0.326***
					(0.037)
Perceived threat to freedom					-0.030
					(0.033)
Personal costs					-0.164***
					(0.026)
Personal benefits					0.303***
					(0.023)
Age					0.019***
					(0.002)
Female					0.216***
					(0.052)
Observations	1,998	1,998	1,998	1,998	1,998
R^2	0.066	0.024	0.066	0.477	0.375

* $p<0.05$, **$p<.01$, ***$p<0.001$

---

[29] We note that the coefficients on perceived inauthenticity and exaggeration are greatly attenuated once we add additional covariates (Model (5) of Tables A3-21 and A3-22). We speculate that this could be because perceived inauthenticity and exaggeration are antecedents to some of the mechanisms captured by the additional covariates. For instance, perceiving that the firm is exaggerating the stated impact may *cause* people to infer that the firm does not appreciate customer feedback. That is, feedback appreciation may be the consequence of exaggeration, rather than an independent alternative mechanism. While incorporating these more complex causal relationships into our conceptual model is beyond the scope of the current paper, we acknowledge that such relationships may exist and explain the aforementioned change in coefficients on perceived inauthenticity and exaggeration.

Web Appendix L: Study 2 Replication

Methods

We recruited participants on Amazon's Mechanical Turk. Following our pre-registration, participants were only invited to take the survey if they passed a brief attention check and were using a brand-name computer (rather than having built their own computer). The final sample included 1,505 people (mean age = 41.2, SD age = 13.6; 54.5% female).

Participants were first asked what computer brand they were using. The response options included 10 of the most used computer brands and an "Other" option that allowed them to write-in the brand (done by 4% of participants). Participants then were asked to imagine that they had received help from their respective computer company's customer service the previous day, the customer service had been satisfactory, and all their questions had been promptly answered. Next, participants imagined receiving an invitation from their respective computer company for a customer feedback survey that would take about 2 minutes. We randomly assigned participants to read either the control subject line ("[Company] customer experience survey invitation") or an impact appeal subject line ("Your voice is important: Shape the [Company] customer experience"), which were identical to the email lines used in the field experiment. Every participant's specific computer brand name was piped into the subject line in the place of "[Company]".

Following the manipulation, participants rated the extent to which the company's use of the subject line seemed inauthentic. We used the same two items to measure *perceived inauthenticity* as used in Study 2 (adapted from Campbell 1995). The two items were highly correlated (r = .70) so we averaged them together to create a composite score of perceived inauthenticity. On the following screen, we asked participants about their *survey completion intentions*: "If you were in this situation, how likely would you be to complete the customer experience survey?" (1- Extremely Unlikely; 6 - Extremely Likely). Participants next completed some descriptive questions about their past experience with their computer company, including the number of years they had been a customer, the number of times they had interacted with the company's customer service in the prior year, their general satisfaction with those interactions, whether they had received an email invitation asking for customer feedback in the prior year, and if so, whether they had responded to that invitation.

After reporting their age, gender, education, and income, participants responded to a 6-item scale measuring their general trusting attitudes towards business, adapted from a measure of general interpersonal trust developed by Levine et al. (2018). Example items include "I believe that companies do not intentionally misrepresent their products and services to customers" and "If I purchased an item from a company and it did not work as advertised, I believe the company would offer me a full refund for the item" (1 - Strongly Disagree; 6 - Strongly Agree). The six items were averaged into a single measure of *trust in business* (α = .84). The full set of items are reported later in this appendix. We also included an 8-item scale about *general interpersonal trust* (adapted from Levine et al. 2018). Finally, we asked respondents whether they had ever received an email with a subject line similar to the one in their experimental condition.

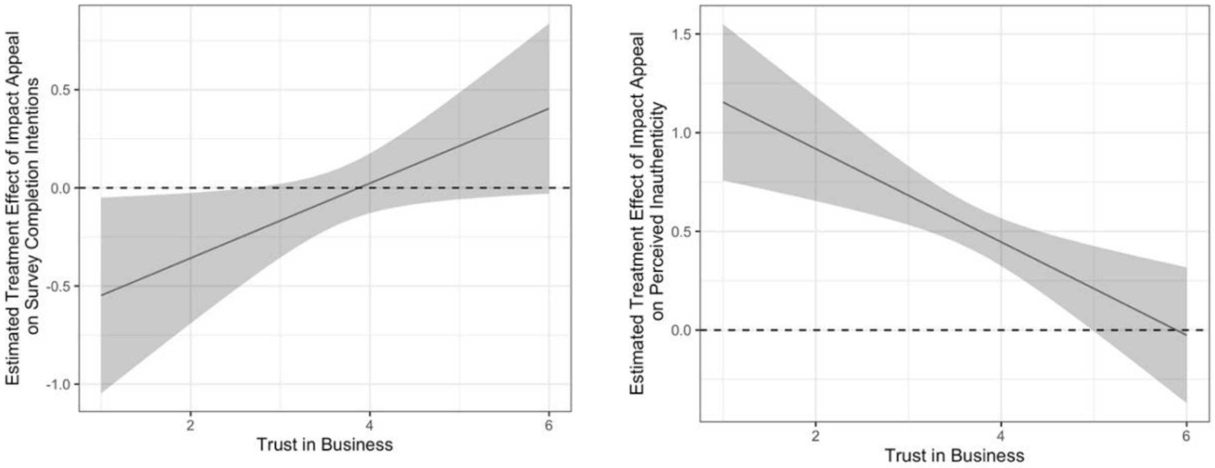
Results

We first tested whether trust in business moderated the effect of the impact appeal on survey completion intentions (Hypothesis 2). We ran an OLS regression that predicts participants' survey completion intentions. The right-hand side includes a dummy variable indicating whether participants were in the Impact (vs. Control) condition, trust in business (mean-centered), and the interaction between the two variables. In line with our field study, we found that the effect of the impact appeal on completion intentions depended on trust in business (interaction b = .190, p = .036; see Table A3-23). Figure A3-1 Panel A depicts the regression-estimated simple effects on completion intentions at every level of trust in business. These results support Hypothesis 2 that the impact appeal was more likely to reduce survey completion intentions among individuals with lower trust in business than among individuals with higher trust in business. To explore whether these findings are unique to trust in business or more generally about trust, we examined whether general interpersonal trust (Levine et al. 2018) also moderated the effect of the impact appeal on survey completion intentions. We did not find evidence of statistically significant moderation (see details at the end of this appendix).

We next tested whether the effect of the impact appeal on perceptions of inauthenticity was moderated by trust in business (Hypothesis 1a). We ran an OLS regression to predict perceived inauthenticity as a function of an indicator for the Impact (vs. Control) condition, trust in business (mean-centered), and their interaction. See Table A3-23. The positive coefficient on the Impact indicator shows that compared to the Control condition, the impact appeal significantly increased perceived inauthenticity at the mean level of trust in business (b = .518, p < .001). Importantly, in support of Hypothesis 1a, the impact appeal increased perceived inauthenticity to a greater extent among people with lower trust in business (interaction b = -.236, p = .002). Panel B of Figure A3-1 displays the simple effects of the impact appeal on perceived inauthenticity at every level of trust in business.

We further tested whether perceived inauthenticity helped explain the more negative effects of the impact appeal on survey completion intentions among people with lower trust in business. In support of Hypothesis 1b, we first confirmed that perceived inauthenticity negatively predicted completion intentions (b = -.451, p < .001). See Table A3-24. Next, we used moderated mediation analyses to test our full theoretical model. Specifically, we tested whether the indirect effect of the impact appeal on completion intentions through perceived inauthenticity depended on trust in business. Using 5,000 bootstrapped samples, we estimated that the impact appeal had a negative indirect effect on completion intentions via perceived inauthenticity among people with moderately low trust in business (who on average indicated "disagree" on the trust in business items; indirect effect = -.550, 95% CI = [-.803, -.356]). Yet for people with moderately high trust in business (who on average indicated "agree" on the trust in business items), the indirect effect of the impact appeal through perceived inauthenticity was not statistically significant (indirect effect = -.066, 95% CI = [-.160, .009). Importantly, the indirect effect was statistically significantly larger for people with moderately low trust in business compared to people with moderately high trust in business (p < .001).

Figure A3-1. Treatment Effects of Impact Appeals Depend on Trust in Business (Study 2 Replication).
This figure shows the regression-estimated simple effects of the impact appeal on survey completion intentions (Panel A) and perceived inauthenticity (Panel B) at every level of trust in business. The grey bands contain the 95% confidence intervals. The trust in business scale points correspond with the following labels: 1 = Strongly Disagree, 2 = Disagree, 3 = Somewhat Disagree, 4 = Somewhat Agree, 5 = Agree, and 6 = Strongly Agree.



Panel A. Survey Completion Intentions

Panel B. Perceived Inauthenticity

As robustness checks, we confirmed that the primary moderation results (Hypotheses 1a and 2) were robust to the inclusion of demographic controls (age, gender, income, and education) and their interactions with the treatment indicator (see Table A3-23). Thus, our observed moderating effects of trust in business do not seem to be explained by demographic differences between participants. We also confirmed that the negative relationship between perceived inauthenticity and completion intentions (Hypothesis 1b) was robust to the inclusion of the demographic controls (see Table A3-24).

Study 2 Replication Supplemental Information

To measure trust in business in this study we used the following prompt.

> The following statements are about your beliefs regarding **for-profit companies** in general.
>
> Please rate the extent to which you agree or disagree with each statement. [1 – Strongly disagree; 6 – Strongly agree]
>
> 1. If a company provided a customer satisfaction guarantee, I believe that the company would honor that guarantee.
> 2. If I purchased an item from a company and it did not work as advertised, I believe the company would offer me a full refund for the item.

3. If a company promised to send me a package and it did not arrive on time, I believe that the company would have a good reason for the delay.
4. If a company knew that a product or service would be bad for my long-term wellbeing, I would not worry that the company might try to sell it to me.
5. I believe that companies do not intentionally misrepresent their products and services to customers.
6. I believe that companies tell customers the truth.

Models 1 and 3 in Table A3-23 report OLS regressions that predict survey completion intentions (Model 1) and perceived inauthenticity (Model 3) as a function of a dummy variable indicating whether participants were in the Impact (vs. Control) condition, trust in business (mean-centered), and their interaction. As a robustness check, we also controlled for standard demographic variables and their interactions with the Impact (vs. Control) condition indicator (i.e., interaction controls) in Models 2 and 4. The demographic variables are: age (mean-centered), female (binary), education (one indicator for whether participants had a college degree or more), and annual household income (one indicator for low income, below $40k, and one indicator for high income, above $80k, with the reference group being the income level between $40k and $80k).

Table A3-23. Moderation by Trust in Business

	Dependent variable:			
	Survey completion intentions		Perceived inauthenticity	
	(1)	(2)	(3)	(4)
Impact	-0.035	-0.046	0.518^{***}	0.294^{*}
	(0.072)	(0.164)	(0.057)	(0.134)
Trust in business	0.417^{***}	0.374^{***}	-0.035	-0.012
	(0.064)	(0.062)	(0.050)	(0.051)
Impact * Trust in business	0.190^{*}	0.186^{*}	-0.236^{**}	-0.223^{**}
	(0.091)	(0.088)	(0.072)	(0.072)
Controls & Interaction Controls	No	Yes	No	Yes
Observations	1,506	1,506	1,506	1,506
R^2	0.081	0.159	0.069	0.096

$^{*}p<0.05$, $^{**}p<.01$, $^{***}p<0.001$

Model 1 in Table A3-24 reports an OLS regression that predicts survey completion intentions as a function of perceived inauthenticity. As a robustness check, we also controlled for standard demographic variables in Model 2: age (mean-centered), female (binary), education (one indicator for whether participants had a college degree or more), and annual household income (one indicator for low income, below $40k, and one indicator for high income, above $80k, with the reference group being the income level between $40k and $80k).

Table A3-24. Perceived Inauthenticity and Survey Completion Intentions

	Dependent variable:	
	Survey completion intentions	
	(1)	(2)
Perceived inauthenticity	-0.451***	-0.403***
	(0.031)	(0.030)
Controls	No	Yes
Observations	1,505	1,505
R^2	0.127	0.191

$^{*}p<0.05$, $^{**}p<.01$, $^{***}p<0.001$

We also asked about general interpersonal trust using the following eight items (adapted from Levine et al. 2018):

The following statements are about your beliefs regarding **other people** in general**.**

Please rate the extent to which you agree or disagree with each statement. [1 – Strongly disagree; 6 – Strongly agree]

1. If someone promised to do me a favor, I believe that the person would follow through.
2. If someone borrowed something of value and returned it broke, I believe the person would offer to pay for the repairs.
3. I would be willing to lend someone almost any amount of money, because I believe that others would pay me back as soon as they could.
4. If someone were going to give me a ride somewhere and the person didn't arrive on time, I would believe there was a good reason for the delay.
5. If someone knew what kinds of things hurt my feelings, I would not worry that the person would use them against me, even if our relationship changed.
6. If I decided to meet someone for lunch, I would be certain the person would be there.
7. I believe that others will not intentionally misrepresent my point of view.
8. I expect that others will tell me the truth.

We averaged the items together to create a single measure of general interpersonal trust (α = .84).

Using the same specification as the one we used for trust in business moderation, we tested whether the effect of the impact appeal on survey completion intentions was similarly moderated by general interpersonal trust. We specifically ran an OLS regression to predict survey completion intentions as a function of an indicator for the Impact (vs. Control) condition, general interpersonal trust (mean-centered), and their interaction. In this case, the coefficient on the interaction term was not statistically significant (b = .158, *p* = .119).

Web Appendix M: Prior Impact Experience Study

This study examines how individuals' prior experience with having an impact moderates the effect of impact appeals. It involves a 2 (Experienced Low Impact vs. Experienced High Impact) X 2 (Impact Appeal vs. Control) between-subjects design.

Methods

Following our pre-registration, we recruited 1,999 participants on Prolific (mean age = 34.8, SD age =13.6, 62.1% female).[1] Participants were asked to imagine themselves as a customer of a large technology company (i.e., "Company X"). In a similar vignette as Study 2, participants imagined that on the previous day they had received help from Company X's customer service, the customer service was satisfactory, and all their questions were promptly answered. Participants then were asked to imagine that today they received an email from Company X requesting their feedback on customer service. They were then asked to imagine that they had offered their feedback to Company X multiple times in the past. They were then randomly assigned to either the "Experienced Low Impact" condition, where they were told that Company X had not made any of the changes they had suggested in the past, or the "Experienced High Impact" condition, where they were told that Company X seems to have made some of the changes they had suggested.

Next, participants were randomly assigned to either the Impact Appeal condition, where they learned that Company X's survey invitation used an impact appeal as the subject line ("Your voice is important: Shape the Company X customer experience"), or the Control condition, where they learned that the survey invitation had a control subject line ("Company X customer experience survey invitation"). Participants then completed questions about perceived inauthenticity (2 items, r = .84), survey taking intentions, exaggeration (2 items, r = .93), ulterior motives (2 items, r = .51), perceived impact (2 items, r = .93), perceived threat to freedom (2 items, r = .69), as well as personal costs and benefits of taking the survey. These scale items were identical to those used in Study 2, except that these items used "Company X" in place of participants' own computer brands. The survey also included an exploratory two-item scale measuring perceptions that Company X values customer impact (r = .90), a two-item measure of *trust in Company X* (r = .87), age, gender, and one item measuring how often they had provided customer feedback over the prior year.

Results

We tested whether prior experience with having had impact on a company moderated the effect of the impact appeal on survey taking intentions. We ran an OLS regression that predicts participants' survey taking intentions. The right-hand side includes a dummy variable for whether participants were in the Impact Appeal (vs. Control) condition, a dummy variable for whether participants were in the Experienced High Impact (vs. Low Impact) condition, and the interaction between these two variables. As pre-registered, in this model and all remaining models in this study, we included a covariate that measured how often participants had provided

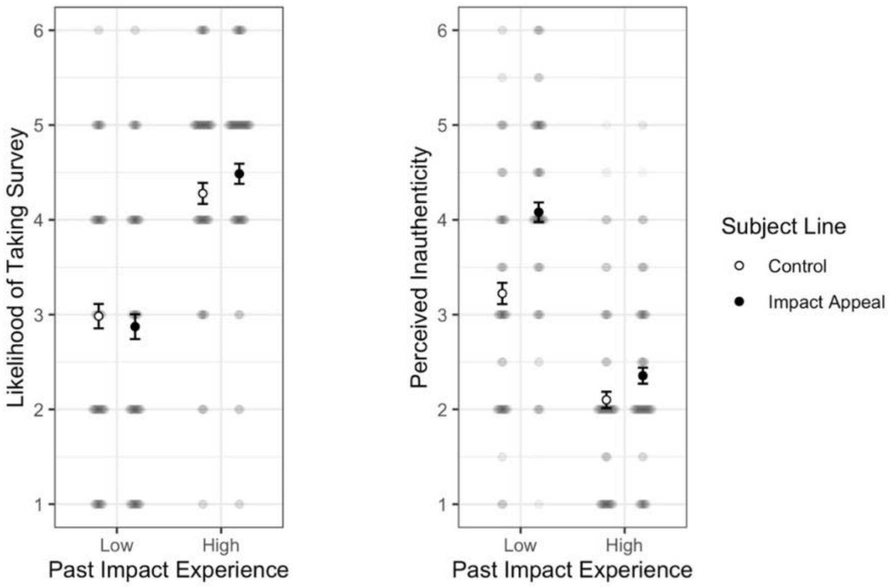
---

[1] In the pre-registration, we originally labelled the Experienced High (vs. Low) Impact manipulation as a manipulation of trust in feedback usage. However, we decided that it was not clearly a trust manipulation and therefore revised the labels here.

customer feedback over the prior year.[2] Randomly assigning participants to imagine experiencing high (vs. low) impact in the past increased the effect of the impact appeal on survey taking intentions (interaction b = .350, *p* = .002). This result is displayed in Figure A3-2, Panel A.

We next tested whether prior experience impacting a company moderated the effect of the impact appeal on perceptions of inauthenticity. We ran an OLS regression to predict perceived inauthenticity as a function of a dummy variable for the Impact Appeal (vs. Control) condition, a dummy variable for the Experienced High Impact (vs. Low Impact) condition, and the interaction between these two variables. Randomly assigning participants to imagine having experienced high (vs. low) impact in the past decreased the effect of the impact appeal on perceived inauthenticity (interaction b = -.607, *p* < .001). This result is displayed in Figure A3-2, Panel B.

Figure A3-2. The Effects of Impact Appeals by Past Experience Making an Impact. This figure shows participants' responses for survey taking intentions (Panel A) and perceived inauthenticity (Panel B) by subject line condition and past impact experience condition. The solid points reflect means, error bars reflect 95% confidence intervals, and translucent points represent each participant's response.



Panel A. Survey Taking Intentions

Panel B. Perceived Inauthenticity

We next tested whether perceived inauthenticity could help explain why having experienced high (vs. low) impact in the past increased the effect of the impact appeal on survey taking intentions. In support of Hypothesis 1b, we again confirmed that perceived inauthenticity

---

[2] Unlike in Study 2, the moderator was manipulated in the hypothetical scenario here, rather than measured. As a result, in this study it is not possible that this covariate (i.e., prior real world experience providing feedback) would be influenced by the moderator (i.e., hypothetical past experience). Thus, unlike in Study 2, we do include this covariate in our models in this study.

negatively predicted survey taking intentions (b = -.571, *p* < .001). Next, we used moderated mediation analyses to estimate whether being randomly assigned to imagine having experienced high (vs. low) impact in the past causally moderated the indirect effect of the impact appeal through perceived inauthenticity on survey taking intentions. Using 5,000 bootstrapped samples, we estimated that, for people randomly assigned to the Experienced Low Impact condition, the impact appeal had a negative indirect effect on survey taking intentions via perceived inauthenticity (indirect effect = -.427, 95% CI = [-.521, -.341]). For people randomly assigned to the Experienced High Impact condition, the indirect effect of the impact appeals through perceived inauthenticity was weaker (indirect effect = -.127, 95% CI = [-.189, -.068]; comparison of indirect effects between the Experienced High Impact vs. Experienced Low Impact conditions, *p* < .001). These results suggest that perceptions of inauthenticity could help explain why the impact appeal increased survey taking intentions to a greater extent when customers imagined having experienced high (vs. low) impact in the past.

As a secondary analysis, we again examined how the effects of impact appeals were explained by perceptions of ulterior motives and exaggerated intentions. We found that randomly assigning people to imagine having experienced high (vs. low) impact decreased the effect of the impact appeal on perceptions of exaggeration (interaction b = -.446, *p* < .001), but it did not moderate the effect of impact appeals on perceptions of ulterior motives (interaction b = -.027, *p* = .769). In univariate regressions, exaggeration (b = -.420, *p* < .001) and ulterior motives (b = -.360, *p* < .001) both predicted weaker survey taking intentions, but when including both of these variables in the same multivariate regression, only exaggeration (b = -.427, *p* < .001) remained a statistically significant predictor (ulterior motives b = .013, *p* = .704). These results suggest that the impact appeal may have had more positive effects on survey taking intentions among people with previous experience making a high (vs. low) impact because these people perceived impact appeals as less exaggerated, not because they perceived weaker ulterior motives.

As a robustness check, we confirmed that moderated mediation models remained significant after controlling for the alternative mediators: perceived impact, perceived threat to freedom, appreciation of impact, personal benefits, and personal costs.

Web Appendix N: Back-of-the-Envelope Calculations

The Impact (Voice) Subject Line vs. the Control Subject Line

In the general discussion of the manuscript, we claim that if our field partner were to implement the Impact (Voice) subject line (rather than the Control subject line) across the seven countries in our study, a back-of-the-envelope calculation suggests that this would result in approximately 35,000 fewer surveys completed in a year.

This is calculated by taking the treatment effect of the Impact (Voice) subject line relative to the Control (b = -.013), multiplying it by the sample size over the 8-week study period (430,666), and then multiplying that by the number of 8-week periods in a year (6.5).

$$\sim 36{,}391 = .013 * 430{,}666 * 6.5$$

Targeted Strategy Using the Full Sample of Customers

We also describe a way that the company can leverage our results to inform a targeted strategy across their customer-base. The general idea is that for each country—including both (a) the seven countries in our field experiment and (b) other countries—we can use our experimental results and data about trust in business to predict which of the five subject lines tested in our field experiment would lead to the highest survey completion rates in that country. For brevity, we call the subject line with the highest predicted completion rate the "optimal subject line" for each country. Then we can project the benefits of using the optimal subject line for each country (i.e., a targeted strategy) in terms of additional completed surveys over a year, as compared to alternative one-size-fit-all strategies. Specifically:

1. For the seven countries in our experiment (which we refer to as the "in-sample countries"), optimal subject lines were selected based on which subject line had the highest completion rate in each country during the experiment.

2. For countries that were not in our experiment, we did not experimentally vary the subject lines so we could not simply select optimal subject lines by observing the best performing subject lines. Instead, the control subject line was used across those countries throughout the duration of our experiment, and we only observed the survey completion rate in response to the control subject line. So, to select optimal subject lines, we needed to predict each of the other four subject lines' expected completion rate in each country. To achieve this objective, we extrapolated from what we know about the countries in our experiment.

    a. Using data from our experiment, we first estimated how trust in business separately moderated the effect of each impact appeal subject line and the time subject line (relative to the Control subject line) on survey completion for the in-sample countries. In an OLS regression, we predicted survey completion as a function of the (mean-centered) country-level measure of trust in business, the four dummy variables separately indicating the three impact appeal subject lines and the time subject line, and the interaction terms between trust in business and the four dummy variables. We included week and country fixed effects and clustered standard errors by country, as in our primary model reported in Table A3-11. Since the addition of covariates did not meaningfully affect the coefficients on interaction terms (see Model (1) vs. (2) in Table A3-25), we continue our back-of-envelop analysis using the output of an OLS regression without additional covariates.

   b. We next identified countries that were not in our field experiment, but were in the field partner's full sample and in the Edelman Trust Barometer data; 21 countries met these criteria, which we refer to as "out-of-sample countries".[3] These 21 countries, together with the 7 in-sample countries, accounted for 89% of our field partner's customer-base.
   c. We then inputted each out-of-sample country's level of trust in business into the regression model we estimated in Step 2.a to calculate the estimated simple effect of each subject line (compared to the Control subject line) for each country. Adding each subject line's estimated simple effect to the observed completion rate in response to the control subject line in each country gave us the predicted completion rate of each subject line.
   d. We then identified the optimal subject line for each country by selecting the subject line predicted to have the highest completion rate. We refer to the completion rates associated with optimal subject lines as "optimal completion rates."

3. We then calculated a population-weighted average of the optimal completion rates to estimate an average optimal completion rate (22.6%) across 7 in-sample countries and 21 out-of-sample countries. To estimate the projected total number of surveys completed if the field partner were to use the optimal subject line in each country, we multiplied the average optimal completion rate (22.6%) by the total number of customers who were invited by our field partner to take the survey over an 8-week period in all 28 countries (593,555 customers), and then multiplied that by the number of 8-week periods in a year (6.5=52/8). That is, if the field partner were to use a targeted strategy and implement the optimal subject line in each country, we estimated the firm would receive 871,932 completed surveys over a year.

$$\sim 871{,}932 = .226 * 593{,}555 * 6.5$$

4. Next, we calculated two relevant benchmarks. We first estimated the projected total number of surveys completed if our field partner were to use the Control subject line (i.e., the status quo before our experiment) across all 28 countries. To do this, we calculated a population weighted average of the countries' completion rates in response to the Control subject lines (21.5%). We then multiplied the average completion rate (21.5%) by the size of the full sample over an 8-week period (593,555), and then multiplied that by the number of 8-week periods in a year (6.5).

---

[3] We could only include countries in this back-of-the-envelope analysis if they had non-missing trust in business scores because trust in business scores are needed to identify the optimal subject lines for countries that were not in our experiment (as explained in steps 2.c and 2.d).

$$\sim 829{,}493 = .215 * 593{,}555 * 6.5$$

Relative to this benchmark of only using the Control subject line, our field partner would receive 42,439 (871,932- 829,493) more completed surveys over a year using the targeted strategy, which reflects a 5% increase.

As a second benchmark, we estimated the projected total number of surveys completed if the field partner were to use the Impact (Voice) subject line across their full sample. To do this, for each in-sample country, we simply used the Impact (Voice) subject line's observed completion rates from the experiment. For each out-of-sample country, we used the predicted completion rates for the Impact (Voice) subject line that we estimated in step 2.c above. We then calculated a population weighted average of the countries' completion rates in response to the Impact (Voice) subject line (20.9%). To estimate the projected total number of surveys completed if our field partner were to only use the Impact (Voice) subject lines, we multiplied the average completion rate (20.9%) by the size of the full sample over an 8-week period (593,555), and then multiplied that by the number of 8-week periods in a year (6.5).

$$\sim 806{,}344 = .209 * 593{,}555 * 6.5$$

Relative to this benchmark of only using the Impact (Voice) subject line, our field partner would receive 65,558 (871,932 - 806,344) more completed surveys over a year using the targeted strategy, which reflects an 8% increase.

Table A3-25. Moderation of Individual Subject Line by Trust in Business (Model Used to Generate the Targeted Strategy)

	Dependent variable:	
	Survey completion	
	(1)	(2)
Time	0.003	0.003
	(0.002)	(0.002)
Impact (Voice)	-0.013***	-0.013***
	(0.002)	(0.002)
Impact (Help)	-0.003	-0.004*
	(0.001)	(0.001)
Impact (Expert)	-0.010	-0.010*
	(0.004)	(0.004)
Trust in business * Time	0.00003	0.00004
	(0.0001)	(0.0001)
Trust in business * Impact (Voice)	0.002***	0.002***
	(0.0002)	(0.0002)
Trust in business * Impact (Help)	0.001***	0.001***
	(0.0002)	(0.0002)
Trust in business * Impact (Expert)	0.002**	0.002**
	(0.0004)	(0.0004)
Controls	No	Yes
Country FEs	Yes	Yes
Week FEs	Yes	Yes
Observations	430,666	430,666
R^2	0.042	0.056

$^{*}p<0.05$, $^{**}p<.01$, $^{***}p<0.001$

Web Appendix O: Process Transparency Study

This study leverages our theory to examine whether the impact appeals can be designed to avoid eliciting negative inferences about inauthenticity. Specifically, in this study, we add a treatment where an impact appeal is accompanied with "process transparency" information about how the firm uses feedback to make impact. It involves a three-condition between-subjects experiment (Control vs. Basic Impact Appeal vs. Transparent Impact Appeal).

Methods

We aimed to recruit 1,500 participants on Prolific. As pre-registered, we only included participant who correctly answered two comprehension check questions, leaving 1,417 participants in our study (mean age = 36.3, SD age =53.4, 42.4% female). Participants first imagined that they were a user of a large social media company (called "Company X"). They were asked to imagine that they received a notification from Company X requesting their feedback about the app. Participants were randomly assigned to imagine receiving one of three notifications:

Control:

> ***Share Your Feedback with Company X***
> *Click to share your feedback in our 2-minute survey.*

Basic Impact Appeal:

> ***Shape the Future of Company X***
> *To have your say in Company X's future direction, click to share your feedback in our 2-minute survey.*

Transparent Impact Appeal:

> ***Shape the Future of Company X***
> *Here is how your feedback can make an impact.*
> *Our product development team will...*
>
> - *review each customer's feedback (within 72 hours)*
> - *decide whether and how to use each customer's feedback (within 3 weeks)*
> - *begin to make changes based on the feedback collected this week (within 2 months)*
>
> *To have your say in Company X's future direction, click to share your feedback in our 2-minute survey.*

Participants then completed questions about perceived inauthenticity, survey taking intentions, perceptions that the firm appreciates customer feedback, and perceptions that users can significantly impact the app. The survey also included a measure of the perceived effort that the company put into crafting the message, age, gender, and one item measuring how often they had provided customer feedback over the prior year.

Results

To estimate the main effects of the impact appeals, we used an OLS regression with two key independent variables: a dummy variable for whether participants were in the Basic Impact Appeal (vs. Control) condition and a dummy variable for whether participants were in the Transparent Impact Appeal (vs. Control) condition. We then used a Wald test to compare the effects of the Transparent (vs. Basic) Impact Appeal conditions. As pre-registered, we included a covariate that measures how often participants had provided customer feedback over the prior year, but the results are robust to excluding this covariate.

We first tested how basic and transparent impact appeals affect survey taking intentions. Relative to the Control message, the Basic Impact Appeal did not significantly increase survey taking intentions (b = -.004, $p = .957$), but the Transparent Impact Appeal did significantly increase survey taking intentions (b = .432, $p < .001$). The Transparent Impact Appeal significantly increased survey taking intentions compared to the Basic Impact Appeal ($p < .001$).

Next, we examined how the impact appeals affected perceived inauthenticity. Relative to the Control message, the Basic Impact Appeal significantly increased perceived inauthenticity (b = .259, $p = .001$), but the Transparent Impact Appeal did not significantly increase perceived inauthenticity (b = .105, $p = .175$). The Transparent Impact Appeal significantly decreased perceived inauthenticity compared to the Basic Impact Appeal ($p = .047$).

We also examined whether perceptions of inauthenticity could help explain differences in survey taking intentions across conditions. Using 5,000 bootstrapped samples, we estimated that, relative to the Control message, the Basic Impact Appeal had a negative indirect effect on survey taking intentions via perceived inauthenticity (indirect effect = -.099, 95% CI = [-.160, -.040]). Importantly, perceived inauthenticity also mediated the difference in survey taking intentions between the Transparent Impact Appeal and the Basic Impact Appeal (indirect effect = .062, 95% CI = [.002, .124]). As a robustness check, we confirmed that mediation models remained significant after controlling for alternative mediators including (1) perceptions that the firm appreciates customer feedback, (2) perceptions that users can significantly impact the app, and (3) perceptions of the effort that the company put into crafting the message.

REFERENCES

Abramson, Paul R. and John H. Aldrich (1982), "The Decline of Electoral Participation in America," *The American Political Science Review*, 76 (3), 502–21.

Amengual, Matthew and Evan P. Apfelbaum (2021), "True Motives: Prosocial and Instrumental Justifications for Behavioral Change in Organizations," *Management Science*, 67 (8), 5032–51.

Angrist, Joshua D. and Jörn-Steffen Pischke (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Bandura, Albert (1977), "Self-Efficacy: Toward a Unifying Theory of Behavioral Change," *Psychological Review*, 84 (2), 191–215.

Bandura, Albert (2001), "Social Cognitive Theory: An Agentic Perspective," *Annual Review of Psychology*, 52 (1), 1–26.

Bandura, Albert (2006), "Toward a Psychology of Human Agency," *Perspectives on Psychological Science*, 1 (2), 164–80.

Barasch, Alixandra, Jonathan Z. Berman, and Deborah A. Small (2016), "When Payment Undermines the Pitch: On the Persuasiveness of Pure Motives in Fund-Raising," *Psychological Science*, 27 (10), 1388–97.

Berry, Leonard L., Lewis P. Carbone, and Stephan H. Haeckel (2002), "Managing the Total Customer Experience," *MIT Sloan Management Review*, 43 (3), 85–89.

Bone, Sterling A., Katherine N. Lemon, Clay M. Voorhees, Katie A. Liljenquist, Paul W. Fombelle, Kristen Bell Detenne, and R. Bruce Money (2017), "'Mere Measurement Plus':

How Solicitation of Open-Ended Positive Feedback Influences Customer Purchase Behavior," *Journal of Marketing Research*, 54 (1), 156–70.

Brehm, Jack W. (1966), "A Theory of Psychological Reactance," Academic Press.

Brown, Christina L. and Aradhna Krishna (2004), "The Skeptical Shopper: A Metacognitive Account for the Effects of Default Options on Choice," *Journal of Consumer Research*, 31 (3), 529–39.

Buell, Ryan W., Ethan Porter, and Michael I. Norton (2021), "Surfacing the Submerged State: Operational Transparency Increases Trust in and Engagement with Government," *Manufacturing & Service Operations Management*, 23 (4), 781-802.

Campbell, Margaret C. (1995), "When Attention-Getting Advertising Tactics Elicit Consumer Inferences of Manipulative Intent: The Importance of Balancing Benefits and Investments," *Journal of Consumer Psychology*, 4 (3), 225–54.

Campbell, Margaret C. and Amna Kirmani (2000), "Consumers' Use of Persuasion Knowledge: The Effects of Accessibility and Cognitive Capacity on Perceptions of an Influence Agent," *Journal of Consumer Research*, 27 (1), 69–83.

Campbell, Margaret and Amna Kirmani (2008), "I Know What You're Doing and Why You're Doing It: The Use of the Persuasion Knowledge Model in Consumer Research," in Hugvstedt, Curt, Herr, Paul and Frank Kardes, eds., *Handbook of Consumer Psychology*, Psychology Press: New York, 549-574.

Chandon, Pierre, Vicki G. Morwitz, and Werner J. Reinartz (2004), "The Short- And Long-Term Effects of Measuring Intent to Repurchase," *Journal of Consumer Research*, 31 (3), 566–72.

Chaudhuri, Arjun and Morris B. Holbrook (2001), “The Chain of Effects from Brand Trust and Brand Affect to Brand Performance: The Role of Brand Loyalty,” *Journal of Marketing*, 65 (2), 81–93.

Chen, Chuansheng, Shin-ying Lee, and Harold W. Stevenson (1995), “Response Style and Cross-Cultural Comparisons of Rating Scales Among East Asian and North American Students,” *Psychological Science*, 6 (3), 170–75.

Clee, Mona A., and Robert A. Wicklund. (1980), “Consumer Behavior and Psychological Reactance," *Journal of Consumer Research,* 6 (4), 389-405.

Cryder, Cynthia E., George Loewenstein, and Richard Scheines (2013), “The Donor is in the Details,” *Organizational Behavior and Human Decision Processes*, 120 (1), 15–23.

Cutright, Keisha M., and Eugenia C. Wu (2023), "In and Out of Control: Personal Control and Consumer Behavior," *Consumer Psychology Review*, 6 (1), 33-51.

Darke, Peter R. and Robin J.B. Ritchie (2007), “The Defensive Consumer: Advertising Deception, Defensive Processing, and Distrust,” *Journal of Marketing Research*, 44 (1), 114–27.

DeCarlo, Thomas E. (2005), “The Effects of Sales Message and Suspicion of Ulterior Motives on Salesperson Evaluation,” *Journal of Consumer Psychology*, 15 (3), 238–49.

Dholakia, Utpal M. (2010), “A Critical Review of Question Behavior Effect Research,” in *Review of Marketing Research*, Vol. 7, Naresh K. Malhotra, ed. Bingley, UK: Emerald Group, 145–97.

Dholakia, Utpal M. and Vicki G. Morwitz (2002), “The Scope and Persistence of Mere-Measurement Effects: Evidence from a Field Study of Customer Satisfaction Measurement,” *Journal of Consumer Research*, 29 (2), 159–67.

Edelman (2020), “Edelman Trust Barometer.”

Edelman (2021), “Edelman Trust Barometer.”

Falbo, Toni. (1977), “Multidimensional Scaling of Power Strategies,” *Journal of Personality and Social Psychology,* 35 (8). 537-547.

Fitzsimons, Gavan J., and Donald R. Lehmann (2004), “Reactance to Recommendations: When Unsolicited Advice Yields Contrary Responses,” *Marketing Science,* 23 (1), 82-94.

Franke, Nikolaus, Martin Schreier, and Ulrike Kaiser (2010), "The “I Designed It Myself” Effect in Mass Customization," *Management Science,* 56 (1), 125-140.

Friestad, Marian and Peter Wright (1994), “The Persuasion Knowledge Model: How People Cope with Persuasion Attempts,” *Journal of Consumer Research*, 21 (1), 1-31.

Fuchs, Christoph and Martin Schreier (2011), "Customer Empowerment in New Product Development." *Journal of Product Innovation Management*, 28 (1), 17-32.

Fuchs, Christoph, Emanuela Prandelli, and Martin Schreier (2010), "The Psychological Effects of Empowerment Strategies on Consumers’ Product Demand." *Journal of Marketing,* 74 (1), 65-79.

Füller, Johann, Hans Mühlbacher, Kurt Matzler, and Gregor Jawecki (2009), "Consumer Empowerment Through Internet-Based Co-Creation," *Journal of Management Information Systems*, 26 (3), 71-102.

Goldberg, Marvin E. and Jon Hartwick (1990), “The Effects of Advertiser Reputation and Extremity of Advertising Claim on Advertising Effectiveness,” *Journal of Consumer Research*, 17 (2), 172.

Hofstede, Geert, Gert Jan Hofstede, and Michael Minkov (2010), *Cultures and Organizations: Software of the Mind*, 3rd ed. McGraw-Hill.

Holden, Gary (1992), “The Relationship of Self-Efficacy Appraisals to Subsequent Health Related Outcomes,” *Social Work in Health Care*, 16 (1), 53–93.

Irmak, Caglar, Mitchel R. Murdock, and Vamsi K. Kanuri, (2020) "When Consumption Regulations Backfire: The Role of Political Ideology," *Journal of Marketing Research,* 57 (5), 966-984.

Jain, Shailendra Pratap and Steven S. Posavac (2004), “Valenced Comparisons,” *Journal of Marketing Research*, 41 (1), 46–58.

Kassirer, Samantha, Emma E. Levine, and Celia Gaertig (2020), "Decisional Autonomy Undermines Advisees’ Judgments of Experts in Medicine and in Life," *Proceedings of the National Academy of Sciences,* 117 (21), 11368-11378.

Kim, Tami, Leslie K. John, Todd Rogers, and Michael I. Norton (2019), “Procedural Justice and the Risks of Consumer Voting,” *Management Science*, 65 (11), 5234–51.

Kirmani, Amna and Margaret C. Campbell (2004), “Goal Seeker and Persuasion Sentry: How Consumer Targets Respond to Interpersonal Marketing Persuasion,” *Journal of Consumer Research*, 31 (3), 573–82.

Kirmani, Amna and Rui Zhu (2007), “Vigilant Against Manipulation: The Effect of Regulatory Focus on the use of Persuasion Knowledge,” *Journal of Marketing Research*, 44 (4), 688–701.

Langer, Ellen J. (1975), “The Illusion of Control,” *Journal of Personality and Social Psychology*, 32 (2), 311–328.

Lemon, Katherine N. and Peter C. Verhoef (2016), “Understanding Customer Experience Throughout the Customer Journey,” *Journal of Marketing*, 80 (6), 69–96.

Levine, Emma E, T Bradford Bitterly, Taya R. Cohen, and Maurice E Schweitzer (2018), “Who is Trustworthy? Predicting Trustworthy Intentions and Behavior.,” *Journal of Personality and Social Psychology*, 115 (3), 468–94.

Liu, Wendy and David Gal (2011), “Bringing us Together or Driving us Apart: The Effect of Soliciting Consumer Input on Consumers’ Propensity to Transact with an Organization,” *Journal of Consumer Research*, 38 (2), 242–59.

Lynch, John G. and Stijn M.J. van Osselaer (2022), “Two Types of Theoretical Contributions in Consumer Research: Construct-to-Construct versus Phenomenon-to-Construct Mapping,” SSRN (December 27), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4311119

Mayer, Roger C., James H. Davis, and F. David Schoorman (1995), “An Integrative Model of Organizational Trust,” *Academy of Management Review*, 20 (3), 709–34.

Meyer, Christopher and André Schwager (2011), “Understanding Customer Experience,” *Harvard Business Review*, 85, 116.

Morgan, Neil A., Eugene W. Anderson, and Vikas Mittal (2005), “Understanding Firms’ Customer Satisfaction Information Usage.” *Journal of Marketing*, 69 (3), 131-151.

Morrison, Elizabeth Wolfe and Robert J. Bies (1991), “Impression Management in the Feedback-Seeking Process: A Literature Review and Research Agenda,” *Academy of Management Review*, 16 (3), 522.

Morwitz, Vicki G., Eric Johnson, and David Schmittlein (1993), “Does Measuring Intent Change Behavior?,” *Journal of Consumer Research*, 20 (1), 46.

Multon, Karen D., Steven D. Brown, and Robert W. Lent (1991), “Relation of Self-Efficacy Beliefs to Academic Outcomes: A Meta-Analytic Investigation,” *Journal of Counseling Psychology*, 38 (1), 30–38.

Priester, Joseph R. and Richard E. Petty (1995), “Source Attributions and Persuasion: Perceived Honesty as a Determinant of Message Scrutiny,” *Personality and Social Psychology Bulletin*, 21 (6), 637–54.

Rajavi, Koushyar, Tarun Kushwaha, and Jan-Benedict E. M. Steenkamp (2019), "In Brands We Trust? A Multicategory, Multicountry Investigation of Sensitivity of Consumers’ Trust in Brands to Marketing-Mix Activities," *Journal of Consumer Research,* 46 (4), 651-670.

Sahni, Navdeep S., S. Christian Wheeler, and Pradeep Chintagunta (2018), “Personalization in Email Marketing: The Role of Noninformative Advertising Content,” *Marketing Science*, 37 (2), 236–58.

Sharma, Eesha and Vicki G. Morwitz (2016), “Saving the Masses: The Impact of Perceived Efficacy on Charitable Giving to Single vs. Multiple Beneficiaries,” *Organizational Behavior and Human Decision Processes*, 135, 45–54.

Silver, Ike, George Newman, and Deborah A. Small (2021), “Inauthenticity Aversion: Moral Reactance toward Tainted Actors, Actions, and Objects,” *Consumer Psychology Review*, 4 (1), 70–82.

Spreitzer, Gretchen M. (1995), "Psychological Empowerment in the Workplace: Dimensions, Measurement, and Validation." *Academy of Management Journal*, 38 (5), 1442-1465.

Stajkovic, Alexander D. and Fred Luthans (1998), “Self-Efficacy and Work-Related Performance: A Meta-Analysis,” *Psychological Bulletin*, 124 (2), 240–61.

Wang, Wenbo, Aradhna Krishna, & Brent McFerran (2017), “Turning off the Lights: Consumers’ Environmental Efforts Depend on Visible Efforts of Firms,” *Journal of Marketing Research*, 54 (3), 478–494.

Wathieu, Luc, Lyle Brenner, Ziv Carmon, Amitava Chattopadhyay, Aimee Drolet, John T. Gourville, Nathan Novemsky, Rebecca K. Ratner, Klaus Wertenbroch, and George Wu (2002), “Consumer Control and Empowerment: A Primer,” *Marketing Letters*, 13 (3), 297–305.

Williams, Patti, Gavan J. Fitzsimons, and Lauren G. Block, (2004) "When Consumers Do Not Recognize “Benign” Intention Questions as Persuasion Attempts," *Journal of Consumer Research,* 31 (3), 540-550.

CHAPTER 4: MAKING SENSE OF DOMINATED OPTIONS

IMPLICATIONS OF DOMINATED OPTIONS FOR TRUST AND CHOICE

Joseph S. Reiff [a]

Jonathan E. Bogard [b]

Eugene M. Caruso [a]

Hal E. Hershfield [a]

[a] University of California, Los Angeles
[b] Washington University in St. Louis

Financial Disclosure: This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1650604. We acknowledge financial support from the UCLA Anderson School of Management and the UCLA Anderson Behavioral Lab.

INTRODUCTION AND THEORY

Imagine wanting to purchase a used car, so you browse the selections of two local dealerships. One dealership offers a standard array of used cars, and the other dealership offers a similar set while also including a few demonstrably low-quality cars at unreasonable prices. Consider how the presence of these overpriced “lemons” might affect your attitude toward the dealership. In this project, we show that the inclusion of a dominated option in a choice set can lead consumers to make negative inferences about the trustworthiness of the marketer offering the options (i.e., the choice architect), which can in turn reduce consumers’ likelihood of selecting *any* option from that choice architect.

Context effects—how the inclusion of additional alternatives affect preferences between a fixed set of options—can result from similarity (Tversky 1972), compromise (Simonson, 1989), and attraction (Huber, Payne and Pluto 1982). The current research focuses on attraction effect choice sets, in which the inclusion of a dominated option is theorized to increase the attractiveness of the dominating alternative (Huber, Payne and Pluto 1982). Context effects have typically been studied under conditions of forced choice (as pointed out by Dhar and Simonson 2003), where participants are required to choose among a particular set of options. When participants are additionally given the possibility of abstaining from making any selection at all (i.e., deferring their choice), which notably increases the decision’s external validity, Dhar and Simonson (2003) found that the inclusion of dominated options reduces deferral by alleviating decision conflict. Yet, recent evidence suggests that the relationship between decision conflict and choice deferral may not be reliable (Evangelidis, Levav, and Simonson 2022). We contribute to this research by identifying one unexplored mechanism that explains when and why dominated options can cause consumers to *avoid purchasing* from certain firms; that is, dominated options can engender distrust in the choice architect.

Supporting this possibility, work on Marketplace Metacognition (Wright 2002) and the Persuasion Knowledge Model (Friestad and Wright 1994) has argued that consumers are aware that marketers are goal-driven agents that deploy different marketing tactics to influence consumer behavior (Kirmani and Campbell 2004). The use of certain persuasion tactics (e.g., placing a cute puppy in an advertisement for cookware; Campbell 1995) can lead consumers to infer that the marketer has manipulative intent, reducing purchasing intentions. More broadly, this research has argued that consumers make inferences about the intent of marketers from the way they design marketing tactics. Applying this idea to context effects, Hamilton (2003) showed that consumers are aware that marketers include dominated options in choice sets as an influence tactic to increase the attractiveness of the dominating alternative, but, even with this awareness, consumers are still susceptible to the attraction effect. In other words, despite inferring the dominated option is a deliberate influence tactic, the inclusion of the dominated option still increases the share of consumers choosing dominating alternative.

Our research shares a similar initial premise as Hamilton (2003)—that choice set composition may leak information about the choice architect's motives. However, we document an entirely different inference from choice set composition with unique behavioral consequences, leading to novel implications for theory and practice. First, we directly measure attitudes toward the choice architect and test whether the inclusion of dominated options in choice sets can lower trust in the company offering the options (Studies 1-3). Second, rather than measuring susceptibility to context effects, we measure another important outcome that may be influenced by choice architect trustworthiness: choices between competing firms. We specifically test whether consumers avoid purchasing from firms that offer dominated options (Studies 2 and 3). Finally, we test an important boundary condition of this phenomenon; we examine whether effects of dominated options are attenuated when consumers have explicit information about the choice architect's trustworthiness (Study 3).

Table 4-1. Summary of Key Results.

	Design	Primary Findings
Study 1 N = 476	2-condition between-subjects design; Imagined receiving either… 2 health insurance options (no dominated option) vs. 3 health insurance options (with dominated option)	Dominated option increases distrust (b = .30, 95% CI = [.03, .56], *p* = .03)
Study 2 N = 490	Within-subject; Imagined choosing between… A company offering 2 investment plans (no dominated option) vs. a company offering 3 investment plans (with dominated option)	People avoid investing with a company that offers a dominated option ($\chi^2 = 110.8$, $p < .001$)
Study 3 N = 791	Mixed design; Same choice as in Study 2; Additional 2-condition between-subjects manipulation where participants imagined receiving either: Explicit information about the trustworthiness of the two companies vs. no explicit trust information	People are more willing to invest in a company that offers a dominated option if they have explicit information about whether the company is trustworthy (b = .10, 95% CI = [.03. .17], *p* = .005)

STUDY 1: DOMINATED OPTIONS ENGENDER DISTRUST

Study 1 shows that the presence of a dominated option in a choice set causes consumers to distrust the choice architect.

Methods

We pre-registered to target 500 participants on Amazon's Mechanical Turk who passed a single screener. The study included two additional attention checks and two comprehension checks. Following our pre-registration, we only included participants in our analysis who passed the attention checks. We report results including all remaining participants. Note that the results are qualitatively unchanged when excluding participants who failed the comprehension checks. The final data includes 476 participants (45% female, Mean age = 41.5, SD age = 12.7).

All participants first imagined receiving an email from a health insurance company ("Medico Enterprises Inc) offering plans. In a two-condition between-subjects design, participants were randomly assigned to see an offer that either included two insurance plans ("Two-Option Condition": Plans A & B) or three insurance plans ("Three-Option Condition": Plans A, B, & C). Plan A included a \$94 premium and \$1500 deductible; Plan B included a \$157 premium and \$1000 deductible; Plan C included a \$165 premium and a deductible that was at least \$1500. Plan C is dominated by the other options in the menu. Note that as an exploratory manipulation, within the Three-Option Condition, we randomized participants to see either a \$1500 or \$1600 deductible; there were not statistically significant differences between these two levels for any of our outcomes, so the remaining analyses collapse across these levels.

Next participants were asked: "If you received these plan options, how much would you distrust Medico Enterprises Inc.?" [1 – No distrust at all; 7 – Completely distrust]. Participants also responded to secondary measures including: four items about their perceptions of the company's competence, one item about their perceptions of whether the insurance company is only motivated by self-interest; one item about skepticism about the insurance company's motives; and one item about how much they believe the insurance company will make it difficult to get reimbursed for covered expenses. The study ended with measures of general dispositional trust in healthcare companies, gender, and age.

Results

As predicted, participants in the Three-Option Condition reported greater distrust (M = 3.71, SD = 1.67) compared to participants in the Two-Option Condition (M = 3.33, SD = 1.52; b = .30, 95% CI = [.03, .56], p = .03). We found further exploratory evidence that, compared to participants in the Two-Option Condition, participants in the Three-Option Condition were more skeptical of the insurance company (b =.35, 95% CI = [.13, .58], p = .003), perceived that the insurance company was less competent (b =-.19, 95% CI = [-.36, -.02], p = .03), and believed that the insurance company was more motivated by its own self-interest (b =.30, 95% CI = [.08, .53], p = .01).

Discussion

Study 1 provided initial support for the hypothesis that dominated options heighten distrust towards the choice architect. The study also provided initial evidence that this effect may be driven by both benevolence- and competence-based trust. Next, we show how these inferences affect consumer choice.

STUDY 2: DOMINATED OPTIONS CAUSE AVOIDANCE

Study 2 examines how dominated options affect consumers' choices between competing firms. The study further tests whether perceptions of distrust drive the effects of dominated options on choice.

Methods

We pre-registered to target 500 participants on Amazon's Mechanical Turk to participate in this study who passed two attention checks. The final analysis includes 490 participants (51% female, Mean age = 39.1, SD age = 12.2).

Participants imagined that they were looking to invest some of their savings for retirement. They imagined that they searched the internet and found two different private financial investment companies offering plans. "Investment Company A" was offering two plans (Plans X & Y) and "Investment Company B" was offering three plans (Plans X, Y, and Z). Plan X includes $500 in annual fees and $1/share in transaction fees for each trade; Plan Y includes $1000 in annual fees and $.20/share in transaction fees for each trade; Plan Z includes $1,300 in annual fees and $1.05/share in transaction fees for each trade. Plan Z is dominated by X and Y.

Participants then rated each of the companies on two dimensions: trust and fairness. Specifically, they rated how much they trust the investment company [1-Strongly distrust; 7-Strongly trust] and how fair they perceived the investment company's offering [1- Completely unfair; 7-Completely fair]. They were further given the binary choice: "If you were going to choose one of the companies to invest your money with, which company would you choose?". We also asked participants which plan they would choose from that company. Participants then completed a brief thought-listing task, listing the reasons why they chose to invest with their selected company. At the very end of the survey, participants completed a measure of general dispositional trust, age, and gender.

Results

Our primary pre-registered analysis tested whether participants were less likely to choose to invest with the company offering three plans (including a dominated plan). 26% of participants chose the choice architect offering the dominated option, which is significantly lower than 50% ($\chi^2 = 110.8$, $p <.001$). See the left panel of Figure 4-1.

Conceptually replicating the results of Study 1, we also found that participants trusted the company offering the dominated option less (M = 3.46, SD = 1.58) than the company that did not offer the dominated option (M = 4.36, SD = 1.55; paired t-test, t(489) =12.76, $p < .001$). We found a similar effect on perceptions of fairness. See the middle and right panel of Figure 4-1.

Figure 4-1. The Effect of Dominated Options on Choice, Trust, and Fairness (Study 2). In the left panel, the figure shows the share of people choosing to invest in the company that does not offer dominated options (light gray; 74%) vs. the company that offers dominated options (dark gray; 26%). The middle and right panels show trust and fairness ratings, respectively, of the company not offering dominated options (light gray points reflect individual responses) and the company offering dominated options (dark gray points reflect individual responses). Note that the columns of points get wider and are jittered to show the density of the distributions at each scale point. The single black points reflect means and the error bars reflect 95% confidence intervals.

Next, we tested whether perceptions of trust predicted participants' choices. We constructed "trust lost" as our key independent variable, which reflects the difference between the trust rating for the company offering the dominated option and the company offering no dominated options. We found that the less participants trusted the company offering the dominated option (relative to the company offering no dominated option), the less likely they were to invest with the company offering the dominated option (b = .12, 95% CI = [.10, ,14], $p < .001$). The results for perceptions of fairness mirrored the trust results.

Discussion

Study 2 showed that the presence of dominated options in choice sets can have important consequences for consumer choice. Consumers were less likely to purchase from a company that offered a dominated option than a company that offered an otherwise equivalent choice set. We again found that consumers trusted the company that offered the dominated option less and viewed its offer as less fair, than the company that offered the otherwise equivalent choice set. Finally, we confirmed that when people perceive that offering a dominated option signals the company is not trustworthy, people are less likely to purchase from that company.

STUDY 3: MODERATION BY TRUSTWORTHINESS OF THE CHOICE ARCHITECT

In Studies 1 and 2, we theorized and found evidence that consumers use signals from the composition of choice sets to infer choice architects' trustworthiness, which may help explain how consumers choose between competing firms. In Study 3, we test an important boundary condition of this phenomenon: When consumers have explicit information about the trustworthiness of the choice architect, they may be less likely to interpret the presence of a dominated option as a signal of low trustworthiness. If true, under these conditions, the dominated option would become an irrelevant factor for consumer choice. Specifically, we predict that including explicit information about the choice architect's trustworthiness would attenuate the effect of dominated options on choice.

Methods

We pre-registered to target 800 participants on Amazon's Mechanical Turk to participate in this study who passed a screener and attention check. The final analysis includes 791 participants in a 2-condition between-subjects design (53% female, Mean age = 40.0 SD age = 12.9).

The design used the same vignette and plan attributes as in Study 2. Participants imagined making an investment decision and choosing between two competing companies: Company A offering 2 investment plans (X and Y) and Company B offering 3 investment plans (X, Y, and Z, where Z is dominated by the other two plans).

In this study, however, participants were also randomly assigned to either receive explicit information on the trustworthiness of the choice architect ("Trust Information Condition") or not ("No Trust Information Condition"). In the Trust Information Condition, participants were shown that both companies received a trust rating of 4.87/5.0 from a market research firm's survey of recent customers. Participants in the No Trust Information Condition did not receive this information.

Participants were then asked our primary outcome measure: "If you were going to choose one of the companies to invest your money with, which company would you choose?". Then participants rated both companies on the company's trustworthiness and the fairness of their offers. At the very end of the survey, participants completed a measure of general dispositional trust, age, and gender.

Results

First, we directly replicated the primary result from Study 2. In the No Trust Information Provided Condition, participants were less likely to choose the company offering a dominated option; 37% chose to invest with the company offering the dominated option ($\chi^2 = 24.6$, $p <.001$). However, in the Trust Information Provided Condition, 47% of participants choose to invest with the company offering the dominated option, which is not statistically significantly different from 50% ($\chi^2 = 1.1$, $p = .147$). Indeed, as predicted, the trust information increased consumers' willingness to invest with the company offering the dominated option (b = .10, 95% CI = [.03. .17], $p = .005$).

Discussion

Study 3 highlights a boundary condition of our theory about dominated options. Dominated options only decrease the likelihood that people purchase from companies *when* explicit information about the trustworthiness of the choice architect is not provided. Put differently, consumers seem to make inferences about choice architect trustworthiness only in the absence of other, more direct signals about trust. When consumers already know whether the choice architect is credible, dominated options no longer provide useful, missing information and thus do not influence consumer choice.

GENERAL DISCUSSION

In this work, we show that the inclusion of dominated options in choice sets can reduce trust in the choice architect and can cause consumers to take their business elsewhere. Moreover, we show that dominated options only affect consumer choice when there is not explicit information about the choice architect's trustworthiness.

Our work makes several theoretical contributions. While previous work has not found evidence that context effects depend on consumers' inferences about marketers' motives

(Hamilton 2003), we identify a novel inference that consumers draw from dominated options that has important consequences for consumer choice. We further contribute to the work on context effects in non-forced choice paradigms. Dhar and Simonson (2003) originally claimed that dominated options *reduce* choice deferral, which violates a key assumption of rational choice theory—the assumption of independent irrelevant alternatives (IIA). Meanwhile, recent evidence suggests this finding may not be robust (Evangelidis, Levav, and Simonson 2022). We contribute to this literature by providing evidence on when and why the presence of a dominated option in a choice set can cause consumers to avoid purchasing from the company making the offer. This too is a violation of IIA but in the opposite direction originally proposed by Dhar and Simonson (2003). More generally, our work argues that context effects are not simply driven by cognitive factors. Rather, social inferences can mediate the effects of choice set composition on consequential consumer attitudes and decisions.

In practice, firms often include dominated options in choice sets (e.g., Bhargava, Loewenstein, & Sydnor, 2017). Our research cautions against this approach as it may reduce consumer trust and, as a result, reduce consumer loyalty over time.

REFERENCES

Bhargava, Saurabh, George Loewenstein, and Justin Sydnor (2017), "Choose to Lose: Health Plan Choices from a Menu with Dominated Options," *The Quarterly Journal of Economics,* 132 (3), 1319-1372.

Campbell, Margaret C. (1995), "When Attention-Getting Advertising Tactics Elicit Consumer Inferences of Manipulative Intent: The Importance of Balancing Benefits and Investments," *Journal of Consumer Psychology*, 4 (3), 225–54.

Dhar, Ravi, and Itamar Simonson (2003), "The Effect of Forced Choice on Choice." *Journal of Marketing Research,* 40 (2), 146-160.

Evangelidis, Ioannis, Jonathan Levav, and Itamar Simonson (2022), "A Reexamination of the Impact of Decision Conflict on Choice Deferral." *Management Science*.

Friestad, Marian and Peter Wright (1994), "The Persuasion Knowledge Model: How People Cope with Persuasion Attempts," *Journal of Consumer Research*, 21 (1), 1-31.

Hamilton, Rebecca W. (2003), "Why Do People Suggest What They Do Not Want? Using Context Effects to Influence Others' Choices." *Journal of Consumer Research,* 29 (4), 492-506.

Huber, Joel, John W. Payne, and Christopher Puto. (1982) "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis." *Journal of Consumer Research,* 9 (1), 90-98.

Kirmani, Amna and Margaret C. Campbell (2004), "Goal Seeker and Persuasion Sentry: How Consumer Targets Respond to Interpersonal Marketing Persuasion," *Journal of Consumer Research*, 31 (3), 573–82.

Simonson, Itamar. (1989) "Choice Based on Reasons: The Case of Attraction and Compromise Effects." *Journal of Consumer Research,* 16 (2), 158-174.

Tversky, Amos. (1972), "Elimination by Aspects: A Theory of Choice." *Psychological Review,* 79 (4), 281.

Wright, Peter. (2002), "Marketplace Metacognition and Social Intelligence." *Journal of Consumer Research,* 28 (4), 677-682.

CONCLUSION

My dissertation proposes that the effects of behavioral policies depend on the inferences that individuals draw about the policymakers implementing the interventions. I provide evidence supporting this idea across four distinct contexts: inferences about the urgency of the recommendations of policymakers who offer pre-commitment (Chapter 1), inferences about the support of policymakers who use peer comparison information (Chapter 2), inferences about the inauthenticity of policymakers' messages that emphasize consumer impact (Chapter 3), and inferences about the trustworthiness of policymakers from the inclusion of dominated options in choice sets (Chapter 4). The theories I developed around these inferences can help (1) explain when behavioral policies backfire (Chapters 1-4), (2) inform how to redesign behavioral policies so they have positive effects (Chapters 1 and 3), (3) identify moderators to help explain divergent effects of interventions across contexts (Chapters 3), and (4) broaden the outcome variables that we use to comprehensively evaluate the effects of behavioral policies (Chapters 2 and 4). Next, I will discuss several limitations of my dissertation that open interesting future directions.

Limitation and Future Direction #1: Using Confirmatory Field Experiments. In Chapter 1-3, the surprising effects of our field experiments led us to develop theories regarding social inferences that we later tested with lab experiments. In several instances, the theories in my dissertation make clear predictions about when and for whom behavioral policies should have positive effects on behavior. In future research, I plan to not only learn and build theories from initial field experiments, but to also conduct confirmatory field experiments that deductively test predictions of the final theories presented in the papers.

Incorporating lessons learned from my dissertation, going forward I also plan to build social inferences into theories of human behavior prior to collecting any field data. In one early-stage project, for instance, we are attempting to increase savings behavior by intentionally communicating an implicit descriptive social norm (Reiff, Shu, Hershfield, Benartzi 2023). We hypothesize that "when-framing" (When will you save...) will increase savings relative to "whether-framing" (Will you save...) because the "when-framing" implicitly communicates the policymaker's beliefs that *other people typically save for retirement*. Multiple pilots and a large pre-registered study confirmed our hypothesis, and we are now seeking field partners to test the theory. In future work, more broadly, I plan to continue conducting confirmatory field tests to better assess whether theories that incorporate social inferences can improve the field's ability to accurately make *a priori* predictions about human behavior in the field.

Limitation and Future Direction #2: Interpreting Responses to Hypotheticals. In Chapters 1, 3, and 4, I relied heavily on how people said they *would respond* to hypothetical behavioral policies. This method is arguably necessary to estimate the extent to which psychological mechanisms explain the effects of behavioral policies. However, we now know from other research that lab experiments (with hypothetical behavior) and natural field experiments may differ in important ways, even when the sample characteristics are relatively similar. Empirically, there are two categories of evidence to consider here: direct evidence from lab experiments and indirect evidence from prediction studies. In one prominent example of direct evidence, a lab experiment using hypothetical decisions led to *opposite* conclusions about the relative effectiveness of two behavioral policies, compared to a large-scale natural field experiment assessing the same policies (Dai et al. 2021). There is further evidence that people often do not correctly predict the relative effectiveness of behavioral policies and tend to overestimate the magnitude of effects (Milkman et al. 2021; Otis and De Vaan 2023; See also Chapters 1 and 3).[1] Discrepancies between hypothetical and field data are also broadly consistent with psychological theory; people have difficulty forecasting their future preferences (e.g., Quoidbach, Gilbert, and Wilson 2013) and struggle simulating their *current* preferences in hypothetical situations that are foreign to their current situation (Frederick 2019; e.g., asking a low-income MTurker to "imagine you are looking for a retirement plan…"). Compounding the problem is that behavioral policies are implemented within rich and diverse social contexts, and treatment effects may be moderated by the social contexts in which interventions are implemented (Gallus et al. 2021). Thus, if hypothetical experiments are to accurately reproduce field experiments, participants must be able to accurately simulate their internal state, their social context, and how the two interact when they respond to behavioral policies. Existing theory and evidence suggest that participants may struggle to do this.

All to say, we should be cautious when interpreting estimates from hypothetical experiments. Researchers do not have the necessary tools to assess whether results from a hypothetical study reflect (a) true behavior or (b) patterns of responding introduced solely by the study's hypothetical nature. This is problematic because researchers may only attribute their results to the hypothetical nature of the study when the results do not conform to the researchers' expectations (see Dai et al. 2021 for an exception). In future research, I plan to investigate systematic ways in which hypothetical and field data differ so researchers can better understand when responses to hypothetical questions accurately reflect behavior and when they do not.

---

[1] This is not direct evidence for hypothetical and field discrepancy because people could be basing their predictions on others' behavior, which diverges from their predictions about their own behavior. However, I still interpret these results as indirect evidence supporting the idea that people have generally poor intuition about how people respond to behavioral policies.

Limitation and Future Direction #3: Building an Integrative Framework. The inferences I included in my dissertation represent a convenient set: they are the inferences that I identified to explain the effects of the behavioral policies that I had the opportunity to evaluate. In a new conceptual project (Reiff, Bogard, and Dai 2023), we are attempting to create a framework that integrates findings from across the field to identify a comprehensive set of inferences that individuals draw from behavioral policies. We are specifically interested in inferences that hinder the effectiveness of policies. The project aims to advance both theory and practice.

From a theory building perspective, we are attempting to identify common antecedents that cause individuals to draw social inferences from policy design. The goal is to build testable predictions about when certain inferences will have strong effects on policy responses, and when they will not. By reviewing and integrating findings from across the field, we hope to identify gaps in existing evidence and set an agenda for future research in this domain.

For policymakers, we hope to build a new tool that improves the behavioral design process (Datta and Mullainathan 2014). When policymakers attempt to design an intervention, popular frameworks (e.g., EAST; Behavioural Insights Team 2014) are often used to answer the question: *what positive mechanisms should I leverage?* However, there are not existing frameworks to help policymakers answer a complementary question: *what negative mechanisms should I guard against?* This asymmetry is broadly consistent with the human tendency to focus on additive, rather than subtractive, changes when designing products and systems (Adams et al. 2021). Our framework seeks to remedy this shortcoming in the behavioral design process. The set of inferences proposed in our framework can serve as an audit checklist. As policymakers design their interventions, they can use this checklist to audit their intervention and scrutinize whether the target population will make one of these inferences. Then, they can redesign their intervention to guard against these inferences with the goal of increasing the likelihood that the intervention works as intended in the field.

Taken together, my dissertation presents a broad hypothesis: The effects of behavioral policies depend on inferences that individuals draw about the policymakers implementing the interventions. I provide evidence in support of this hypothesis across four chapters, delineate project-specific and overarching contributions to the theory and practice of behavioral science, and outline limitations and future directions.

REFERENCES

Adams, Gabrielle S., Benjamin A. Converse, Andrew H. Hales, and Leidy E. Klotz (2021), "People systematically overlook subtractive changes." *Nature,* 592 (7853). 258-261.

Behavioural Insights Team (2014) "EAST: Four simple ways to apply behavioural insights." Behavioural Insight Team, London.

Dai, Hengchen, Silvia Saccardo, Maria A. Han, Lily Roh, Naveen Raja, Sitaram Vangala, Hardikkumar Modi, Shital Pandya, Michael Sloyan, and Daniel M. Croymans (2021), “Behavioural nudges increase COVID-19 vaccinations,” *Nature*, 597 (7876), 404–9.

Datta, Saugato, and Sendhil Mullainathan (2014) "Behavioral design: a new approach to development policy." *Review of Income and Wealth,* 60 (1): 7-35.

Frederick, Shane (2019), “Reports, predictions, and extrospections.”

Gallus, Jana, Joseph Reiff, Emir Kamenica, and Alan Page Fiske (2021), “Relational incentives theory,” *Psychological Review*.

Milkman, Katherine L., Dena Gromet, Hung Ho, Joseph S. Kay, Timothy W. Lee, Pepi Pandiloski, Yeji Park, Aneesh Rai, Max Bazerman, John Beshears, Lauri Bonacorsi, Colin Camerer, Edward Chang, Gretchen Chapman, Robert Cialdini, Hengchen Dai, Lauren Eskreis-Winkler, Ayelet Fishbach, James J. Gross, Samantha Horn, Alexa Hubbard, Steven J. Jones, Dean Karlan, Tim Kautz, Erika Kirgios, Joowon Klusowski, Ariella Kristal, Rahul Ladhania, George Loewenstein, Jens Ludwig, Barbara Mellers, Sendhil Mullainathan, Silvia Saccardo, Jann Spiess, Gaurav Suri, Joachim H. Talloen, Jamie Taxer, Yaacov Trope, Lyle Ungar, Kevin G. Volpp, Ashley Whillans, Jonathan Zinman, and Angela L. Duckworth (2021), “Megastudies improve the impact of applied behavioural science,” *Nature*, 600 (7889), 478–83.

Otis, Nicholas G. & Mathjis de Vaan (2023), “Evaluating Managerial Expectations.” Working Paper.

Reiff, Joseph S., Jonathan E. Bogard, & Hengchen Dai. (2023), “Social inferences from behavioral policy design.”

Reiff, Joseph S., Steve Shu, Hal E. Hershfield, & Shlomo Benartzi. (2023), “Using when- (vs. whether-) framing to increase saving behavior.”

Quoidbach, Jordi, Daniel T. Gilbert, and Timothy D. Wilson (2013), “The end of history illusion,” *Science*, 339 (6115), 96–98.