

Chapter 6: Why Many Behavioral Interventions Have Unpredictable Effects in the Wild:

The Conflicting Consequences Problem

Indranil Goswami and Oleg Urminsky

Forthcoming in *Behavioral Science in the Wild*, Nina Mazar and Dilip Soman, Eds

Changing the behavior of a well-functioning car is straightforward. Pressing the gas pedal has one simple effect, accelerating the car, while pressing the brake has the opposite effect, decelerating the car. Changing the behavior of the driver is a different matter. Pushing a behavioral “pedal” to induce safer driving won’t necessarily have a single predictable effect. Drivers wearing a seat belt are more likely to survive a crash. So, it might seem straightforward that mandating seat belt usage should simply save lives. However, mandating seat belt usage can have the additional effect of more careless driving, partially offsetting the beneficial effects.¹

More recently, similar concerns have resulted in much debate about attempts to mitigate the COVID pandemic by encouraging facemask wearing.² Concerns that requiring people to wear masks might make them less careful about following social distancing norms or might increase face-touching played a role in the initial guidance against mask-wearing in the United States.³ In this case, the concern was overblown, as mask-wearing is effective⁴ and does not seem to reduce social distancing.⁵ However, multiple studies on “risk compensation” across various domains⁶ find a wide range of results, illustrating the difficulty of knowing the effect of a particular intervention in advance.

We provide a framework for understanding why promising behavioral interventions can have inconsistent and unpredictable outcomes. We describe how the net effect of a single intervention often results from a balance between multiple potentially conflicting psychological

consequences of the intervention. The framework can help practitioners make sense of the differences they observe in the results of nudges across field implementations, as well as to be better calibrated regarding the uncertainty inherent in behavior change tactics. We argue that behavioral interventions validated in one setting can seldom be taken “off the shelf” (see Chapter 1) and successfully applied to a different context, and we encourage practitioners to use in-context field experiments to determine how the various consequences of a behavioral intervention net out in their given situation.

The Problem of Multiple Conflicting Consequences

As popularized in the idea of “nudges,” behavioral interventions may seem simple, as if intervening on people’s decision process would shift their decisions in one particular direction. However, that seeming simplicity may be deceptive. Consider the example of descriptive social norms – a widely-used nudge based on the idea that people are motivated to match their peers’ desirable behaviors. Accordingly, people who are doing less of a desirable behavior can be influenced to do more by showing them information about their peers’ behavior. Interventions using such social norms have been shown to be effective in the areas of tax compliance,⁷ recycling,⁸ energy conservation,⁹ water conservation,¹⁰ and other prosocial behaviors. However, in practice, such interventions are not equally effective across implementations. Identifying the forces at work in such interventions can help us understand why and how the effectiveness of descriptive norm interventions varies across contexts and populations.

For example, Opower leverages the principle of descriptive social norms by providing consumers with information about how their energy consumption compares with that of their neighbors to motivate the conservation of electricity.¹¹ According to the theory underlying the

descriptive norm intervention, presenting norm information (e.g., the average amount of electricity used) emphasizes that many people conserve and use a modest amount of electricity, motivating consumers to begin or continue to engage in the same desired behavior. However, the information can also lead to or strengthen a belief that some people do not engage in the behavior (i.e., that some fail to conserve electricity), which can reduce the motivation to engage in the behavior, giving people the “license” to likewise fail to conserve.¹²

Whether the intervention is effective overall then depends on the relative strength of these conflicting consequences. If a person reading the information focuses more on the other people who do engage in the desired behavior and is primarily motivated to join them, then the intervention will increase the behavior (e.g., energy conservation) for that person. If, on the other hand, the person reading the information also focuses on those other people who do not engage in the behavior and is equally demotivated by these people, the two consequences may cancel, resulting in no observed effect of the intervention for that person. In fact, for a person who primarily focuses on those who don’t engage in the behavior and is demotivated (or licensed) by this information, the intervention could even backfire, reducing the person’s energy conservation. After all, few people want to be the sucker who is engaging in costly prosocial behavior that seems pointless because others take the easy way out and avoid doing their part.

Different Consequences Among Different People and in Different Contexts

On its own, the problem of multiple consequences does not have to be a major impediment to reliably applying behavioral science. To know whether an intervention with multiple conflicting consequences will be successful overall, we simply need to know whether the positive effects are larger than the negative ones. If a rigorous experiment finds positive

effects of the intervention (e.g., providing electricity usage norms reduces usage), then we can conclude that, on average, the positive effects (motivation to conserve from knowing that many others do so) outweigh the negative effects (demotivation from knowing that some others don't). In fact, such an experiment could even help us quantify the relative impact of the two factors.

However, deciding to implement an intervention based on the results of a study in another setting requires assuming that the relative impact of the multiple consequences is stable over populations and contexts. This strong assumption is typically not true. Indeed, several examples in this book in domains ranging from savings behaviour (Chapter 1), credit card debt (Chapter 3), pension contributions (Chapter 3), energy savings (Chapter 1), and welfare programs (Chapter 4) highlight the perils of assuming the stability of settings in translating knowledge.

As illustrated in Figure 6.1, in one application where the positive effect of consequence A outweighs the negative effect of B, a net effect results, while no net effect is observed in another application where the two consequences are equal and cancel each other. Research on psychological factors that influence preferences and choices typically identifies a wide variety of moderators – factors that make a particular psychological process more or less likely to occur and make the process stronger or weaker when it does occur. As a result, the mix of consequences of the same intervention often varies across contexts.

<<INSERT FIGURE 6.1 ABOUT HERE>>

Figure 6.1 The conflicting consequences problem

The same intervention could yield different results in different applications based on the relative strengths of consequences A and B on the desired behavior.

A recent empirical investigation found that Opower's norm-based nudges to promote energy conservation yielded differing net results across implementations. In particular, the intervention yielded significantly higher energy savings in their first ten sites than in their next 101 replications.¹³ In general, the available information about the effectiveness of interventions may not be representative. Organizations developing a new program might target the initial tests for the most promising locations. Organizations with a more successful ongoing program might be more willing to do a program evaluation field experiment, while those who think evaluations could undermine their credibility would be less willing.¹⁴ Successful interventions are also more "newsworthy" and may be more likely to be publicized.¹⁵

To understand how the effects of descriptive norms may differ, we can identify some likely factors that contribute to differences in the relative strength of the multiple consequences of this intervention across people and contexts. Before the intervention, some people used less energy than others. Those who used less energy might have assumed that others were like them, carefully conserving energy. As a result, these conservationists might then be surprised to learn that the average person uses far more energy than they had thought, leading them to focus on those not conserving and to feel demotivated. Among those using more energy, on the other hand, the relative strength of the motivating and demotivating effects is likely to depend on pre-existing attitudes. For those non-conservers who endorse the goal and don't feel threatened by the implicit rebuke the information provides, the motivating effects are likely to dominate, as intended. However, those who do not endorse the goal or who are defensive about their non-normative behavior may avoid the intervention, reducing its efficacy.¹⁶ Even more problematically, they may feel more licensed to not engage in the behavior themselves when

they know that there are others who do not engage in the behavior.¹⁷ Consistent with this possibility, descriptive norms are more effective at reducing electricity consumption among liberals than conservatives.¹⁸

In theory, if an intervention only has a single simple effect, decisions about using the intervention are simpler, even when the magnitude of the effect varies by context. As long as the cost of the intervention is low, implementation can be justified by the potential benefit if successful and the minimal risk otherwise (because nothing will change if it is unsuccessful). However, this justification no longer holds when the intervention has multiple *conflicting* (i.e., both positive and negative) consequences that vary in relative strength. In such situations, the risk is that the negative consequences could dominate in a particular implementation, resulting in a negative net effect.

Our general framework applies to any situation in which a successful intervention from one study is used in a different setting, as illustrated using three case studies.

Case 1: Using Temporary Incentives to Motivate Useful Behavior

Material incentives (including financial rewards, such as salaries) are perhaps the most basic way to motivate behavior because of the simple fact that people will do more of an activity that is more rewarding. However, a highly influential literature has raised concerns about a conflicting consequence when using temporary incentives to motivate useful behavior (e.g., going to the gym, eating healthy, giving nutritional supplements to children in developing countries). Researchers in psychology, based on lab experiments, have argued that although such incentives boost compensated behavior, they also undermine people's own intrinsic motivation to do useful or enjoyable activities without compensation.¹⁹ As a result, practitioners have been warned²⁰ that once rewards are taken away, people's long-term engagement and performance

might be lower than if incentives had not been used. This view has been quite influential, limiting the use of temporary incentives as a tool to motivate behavior, particularly in the context of education.

However, our framework recognizes that a *potential* conflicting effect of reducing intrinsic motivation is not a sufficient reason to abandon the use of temporary incentives. Instead, the question is how the potential motivating and demotivating consequences of temporary incentives balance out to produce the net effect of the intervention on the specific outcome of interest in a given context. In fact, field experiments that measured the effects of temporary incentives on people's behavior days, weeks, or months after the incentive ended reported no such detrimental net effects, contrary to the immediate negative effects found in lab experiments.²¹ For example, temporary cash awards for students raising their first-year GPA had either no effect or a positive effect on GPA the subsequent year.²²

Our research confirms that temporary incentives have two separate conflicting effects on post-incentive motivation and engagement.²³ However, the potential adverse effects are not due to a long-term loss of intrinsic motivation. Instead, decision-makers are motivated to take a "break" after exerting additional effort to earn the reward. When a moderately rewarding temporary incentive is provided, we find a momentary break from task engagement when the reward ends, but the net effect differs over a longer time horizon, as people quickly return to their baseline activity level. A sufficiently rewarding incentive can increase liking of the task because of positive reinforcement from the rewards. Indeed, when the temporary incentive is more generous, we observe less of a momentary decrease in engagement immediately after the incentive ends, and people's engagement can increase to above their pre-reward levels over time.

As a result, the net effect will systematically depend on both the magnitude of the reward and the time-frame of the outcome (e.g., immediately after the incentive ends vs. longer-term).

Identifying the actual conflicting consequences allowed us to reconcile past research and better predict the effects of temporary incentives. We applied this approach in a café field study where we used a discount to incentivize lunchtime customers to buy soup. As predicted, soup sales dropped significantly on the first day after the discount ended (without advance notice) but then increased back to the pre-incentive sales level over the next few days.²⁴

Case 2: Can Defaults Increase Charitable Contributions?

Cass Sunstein, the co-author of the book *Nudge*, has stated that if there were an Olympic medal for the most effective tool in behavioral economics, the clear winner would be decision defaults,²⁵ a view supported by a review comparing behavioral interventions.²⁶ Defaults define one option as the action to take unless the person selects a different course of action. In some settings (e.g., organ donation,²⁷ saving for retirement,²⁸ etc.), a major reason why defaults are effective is that many people don't make a decision and instead let the default occur.

However, defaults can affect outcomes even when people actively make decisions, and a default course of behavior cannot be imposed, such as in fundraising appeals. For example, people might treat the default as an implicit recommendation or norm, or they might be averse to change and prefer the status quo.²⁹ Most philanthropic organizations solicit donations online or via mail, where it is trivial to implement a non-binding default or “suggested amount.” Our review of fundraising practices of top charities found that less than half used a pre-selected default contribution in their menu of amount options. Of those who had a default-ask, the vast majority used the lowest or second-lowest menu amount as the default donation option. We

conducted a series of studies to test the net effect on the funds raised of defaulting potential donors to a particular donation amount.

Prior experimental research on suggested donation amounts had revealed a confusing set of seemingly contradictory results, with suggested amounts sometimes resulting in more donations, sometimes less, and sometimes making no difference. Our investigation revealed that the effect of default nudges on the funds raised by a charitable solicitation could not be understood as a single effect, or even a single effect that varies in strength.³⁰ Instead, in a series of lab and online studies, we found that two separate conflicting consequences consistently operated in tandem, systematically varying with the defaulted donation amount, such that their joint effect determines whether implementing a particular default “suggested donation” raises more funds than the baseline of no default.

First, defaults suggest to the donor an amount to give. So, the straightforward (and typically intended) effect of defaulting or suggesting an amount is that people give that amount or a similar amount. Thus, among those who donate, setting a higher default would be expected to yield larger donations, on average, and we find that it usually does. In fact, this consequence results in a risk that setting too low a default – asking donors for less than they would otherwise be willing to give – could reduce donations. Donors asked to donate \$10, for example, may think to themselves, “I was willing to give \$50, but since you say that all you need is \$10, it’s a win-win: I get to feel good about helping, for less!” If this were the only effect of defaults, the prescription would be clear: set a high donation amount as the default.

However, we find that there is also a second effect of suggested amounts, conflicting with the first: suggesting a donation amount impacts whether or not people donate at all. Even when people can donate any amount they choose, suggesting a large amount typically reduces

donations, and suggesting a small amount increases donations, relative to the baseline. It's as if potential donors would rather not participate than give a "wrong" amount. Donors reassured by the suggestion of a small donation are more likely to open their wallets, while those asked for a large donation move on instead of giving the small amount they would have otherwise been willing to donate.

As a result, the net effect of setting a particular donation amount as the default is not obvious. Asking for a smaller amount than the typical donation will encourage people to donate, but they may give less than they would have given without a default. In contrast, asking for an above-average amount may result in fewer people participating but giving more. Given that the net funds raised are the product of participation and average donation amount, how the joint effect of these two consequences balances out will determine whether a particular suggested amount results in more funds, less funds, or the same amount as without the suggested amount. A re-analysis found that the seemingly contradictory results of past studies were largely explained as the net of these two conflicting consequences varying by the suggested amount.

Furthermore, the relative sensitivity of participation rates and averaged donation amounts to the default intervention may differ across people and donation contexts, resulting in even greater uncertainty about the net effect of an intervention in a given context. In particular, among prior donors who give regularly, participation may be less sensitive, and donation amount may be relatively more sensitive to the default, while the reverse may be the case among donors who often fail to give.

Results of two separate field experiments conducted with the University of Chicago Booth School of Business annual fundraising campaign validated these insights and demonstrated the usefulness of the framework. In the first experiment, conducted at the end of

the annual campaign, we varied the presence and amount of a “suggested” donation level in solicitations sent to prior donors who had not yet given in response to three mailers sent earlier that academic year. Among these unlikely donors, setting the lowest amount as the default raised more funds than the no-default appeal, while the higher default amounts were not effective. The low default did result in lower donation amounts among those who gave, but this was more than compensated for by a 128% increase in the (low) rate of giving.

In a follow-up study, we conducted the same experiment at the start of a new annual campaign in solicitations sent to all past donors, including the donors who give regularly that had been excluded in the first study.³¹ The low default, which had been the best performer in the prior campaign, instead did the worst, reducing funds raised compared to the no-default baseline. This outcome was the result of the net effect from a different balance between the two conflicting outcomes: the low default substantially reduced the donation amounts among those who gave but had only a small positive effect on the (already much higher) participation rate in this population.

The experiment targeting frequent donors found a net effect of the low-amount default that was the *opposite* of the effect in the study targeting infrequent donors. While these findings can provide guidelines for thinking about how suggested amounts might be beneficial or harmful for a charity to use, a particular organization cannot readily know how a particular suggested amount would affect the funds raised. Even when we believe that we understand the underlying process, the variability of the relative impact of conflicting consequences leaves us uncertain and unable to predict the net effect of a specific default intervention when targeting a particular pool of donors for a given charity.

Case 3: The Failure of a Promising Matching Mechanism in Fundraising

It may turn out that the conflicting consequences problem is occurring in practice, even when conflicting consequences were not identified in advance, and the proposed intervention may have seemed like it would have straightforward positive effects. Research that we have conducted regarding matching solicitations in charitable fundraising (e.g., every dollar donated is matched by an extra dollar from a benefactor) provides a demonstration of this principle.³²

Matching appeals have been described as a staple of fundraising and are routinely used in appeals.³³ Theoretical models of altruism identify two benefits to the donor: the charity having more funds to do its work (regardless of the source) and the “warm glow” that the donor gets from being personally responsible for giving.³⁴ In this model, matching appeals motivate donors by enabling the donor to direct more funds to the charity than just the donation they personally give, increasing the impact of the person’s donation on the funds available to the charity, and thereby motivating them to give more.³⁵ Based on this theoretical framework, we developed a novel approach to making matches more effective: reframing the match as the benefactor helping the donor give more (i.e., as opposed to the benefactor making their own separate matching donation). The idea behind this “giving credit” match framing is to increase the “warm glow” the donor feels from giving, in addition to increasing the objective benefit to the charity. Consequently, the new framing would be predicted by theoretical models to increase donations from the match. A sample of fundraising experts who we surveyed overwhelmingly predicted that the “giving credit” framing would be effective, and most believed it would be more effective than the standard match framing.

However, despite its basis in academic theory and the favorable intuitions of fundraising experts, our novel “giving credit” framing was a failure when tested in practice in two fundraising campaigns for a Chicago arts non-profit targeting prior donors. The participation rate among potential donors who received the “giving credit” version of the matching appeal was almost half of the participation rate using the standard matching description and, as a result, raised significantly less funds.

Why did this promising, vetted intervention fail? While our field experiments did not collect the data needed to conclusively identify the underlying cause, it seems clear that the “giving credit” framing had a conflicting negative consequence on donors’ motivations that far outweighed the intended positive effect of giving the donors a reason to feel “warm glow” from the benefactor’s match. It is possible that the “giving credit” framing seemed weird or manipulative to the prospective donors. Alternatively, encouraging the potential donor to see the benefactor’s funds as their own donation may have had the reverse effect, reducing perceived personal accountability for one’s own donation because of the suggested “co-mingling” of funds. In any case, a promising intervention failed in practice, at least in the setting in which it was tested, because the intervention that was assumed to have a single positive motivational impact instead involved multiple conflicting potential consequences, resulting in a net effect mispredicted by us as well as by fundraising experts.

Table 6.1 Summary of the Cases Discussed in This Chapter

	Consequence A	Consequence B
Case 1: Using temporary incentives to motivate useful behavior	Longer-run return to baseline after “break” is over. Positive spillover from reinforcement, habit formation also possible.	Short-run decrease in engagement after incentives end because the option of a “break” is salient and easy to justify
Case 2: Using defaults to increase charitable contributions	Lower defaults motivate participation as donors earn “warm glow” from the act of giving.	Lower default amount licenses people to donate a smaller amount.
Case 3: “Giving credit” match framing intended to improve fundraising outcomes	Increase in “warm glow” from donors’ match.	Unidentified: could be confusion, perceived manipulativeness, or a reduction in accountability.

The “giving credit” framing might not be harmful or could even be successful in a different fundraising context. To know if that is the case, we would need first to identify all the conflicting consequences of framing matching appeals that way and then develop a means to predict how the relative impact of the conflicting consequences varies across implementation contexts. In the absence of this complete understanding of the intervention, a policy-maker simply cannot generalize from observing either success or failure in another context to determine whether an intervention with potentially conflicting consequences would be beneficial, inconsequential, or harmful in their own context.

The Path Forward

The problem of conflicting consequences prescribes a different approach to using behavioral insights in practice than is often the case. Experimental evidence that an intervention has beneficial effects in another context, whether from academic research or observed in practice, validates the intervention as a *hypothesis to be tested* in the intended context, not as a policy to be adopted. Absent a complete understanding of how the potential conflicting consequences net out in a given context, which we typically do not have, we cannot predict the consequences of implementing even seemingly “proven” interventions. For academics, this suggests a need to go beyond “proof of concept” research and develop detailed and robust models of the most promising interventions.

Table 6.2 Decision Aid to Identify Potential Conflicting Consequences: Ask Yourself the Following Five Questions to Start

Factors suggesting multiple consequences	
Question 1	Do people differ in relevant experience (e.g., high vs. low engagement, prior experience, or expertise)?
Question 2	Do people differ in relevant values or attitudes? (e.g., identifying with a pro-environment agenda or not)
Question 3	Can short-term and long-term effects differ (e.g., due to forgetting, fatigue, habit-formation, or satiation over time)?
Question 4	Does the targeted behavior involve multiple distinct decisions (e.g., whether to give and how much to donate)?
Question 5	Do the intervention cues have multiple plausible interpretations? (e.g., as information or allocation of responsibility or an attempt to manipulate)

For practitioners, even the most promising new behavioral intervention should be subjected to substantial field-testing in one’s own context, across the factors varying in that context, before changing to a new policy. The common reluctance many organizations have to test initiatives experimentally with appropriate controls and the preference to instead simply set

policy and act can have serious hidden costs when interventions have conflicting consequences, perpetuating costly mistakes and impeding context-specific informed decision-making.

NOTES

¹ See Peltzman, S. (1975). The Effects of Automobile Safety Regulation. *Journal of Political Economy*, 83(4), 677–725. <https://doi.org/10.1086/260352>; and Cohen, A., & Einav, L. (2003). The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities. *The Review of Economics and Statistics*, 85(4), 828–843. <https://doi.org/10.1162/003465303772815754>.

² Randall, I. (n.d.). *Wearing face masks makes people more careless and less likely to follow social distancing guidelines*. Daily Mail Online. Retrieved March 15, 2021, from <https://www.dailymail.co.uk/sciencetech/article-8659487/Wearing-face-masks-makes-people-careless-likely-follow-social-distancing-guidelines.html>

³ *Fact check: Outdated video of Fauci saying “there’s no reason to be walking around with a mask.”* (n.d.). Reuters. Retrieved March 15, 2021, from <https://www.reuters.com/article/uk-factcheck-fauci-outdated-video-masks/fact-checkoutdated-video-of-fauci-saying-theres-no-reason-to-be-walking-around-with-a-mask-idUSKBN26T2TR>

⁴ Mitze, T., Kosfeld, R., Rode, J., & Wälde, K. (2020). Face masks considerably reduce COVID-19 cases in Germany. *Proceedings of the National Academy of Sciences*, 117(51), 32293–32301. <https://doi.org/10.1073/pnas.2015954117>

⁵ Seres, G., Balleyer, A. H., Cerutti, N., Friedrichsen, J., & Süer, M. (2020). *Face Mask Use and Physical Distancing Before and After Mandatory Masking: Evidence From Public Waiting Lines* (SSRN Scholarly Paper ID 3641367). Social Science Research Network. <https://doi.org/10.2139/ssrn.3641367>

⁶ See Mantzari, E., Rubin, G. J., & Marteau, T. M. (2020). Is risk compensation threatening public health in the covid-19 pandemic? *BMJ*, 370, m2913. <https://doi.org/10.1136/bmj.m2913>; and Esmaeilikia, M., Radun, I., Grzebieta, R., & Olivier, J. (2019). Bicycle helmets and risky behaviour: A systematic review. *Transportation Research Part F: Traffic Psychology and Behaviour*, 60, 299-310. doi:10.1016/j.trf.2018.10.026

⁷ Thaler, R. H. (2012). Watching behavior before writing the rules. *The New York Times*, 7.

⁸ Cialdini, R. B. (2003). Crafting Normative Messages to Protect the Environment. *Current Directions in Psychological Science*, 12(4), 105–109. <https://doi.org/10.1111/1467-8721.01242>

⁹ Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9), 1082–1095. <https://doi.org/10.1016/j.jpubeco.2011.03.003>

¹⁰ Ferraro, P. J., & Price, M. K. (2013). Using Nonpecuniary Strategies to Influence Behavior: Evidence from a Large-Scale Field Experiment. *The Review of Economics and Statistics*, 95(1), 64–73. https://doi.org/10.1162/REST_a_00344

¹¹ Rahim, S. (2010, June 21). *Finding the “Weapons” of Persuasion to Save Energy*. NY Times. <https://archive.nytimes.com/www.nytimes.com/cwire/2010/06/21/21climatewire-finding-the-weapons-of-persuasion-to-save-ene-8137.html?pagewanted=1>

-
- ¹² See Ozaki, T., & Nakayachi, K. (2020). When Descriptive Norms Backfire: Attitudes Induce Undesirable Consequences during Disaster Preparation. *Analyses of Social Issues and Public Policy*, 20(1), 90–117. <https://doi.org/10.1111/asap.12195>; Miller, D. T., & Prentice, D. A. (2016). Changing Norms to Change Behavior. *Annual Review of Psychology*, 67(1), 339–361. <https://doi.org/10.1146/annurev-psych-010814-015013>; and Mollen, S., Rimal, R. N., Ruiters, R. A. C., & Kok, G. (2013). Healthy and unhealthy social norms and food selection. Findings from a field-experiment. *Appetite*, 65, 83–89. <https://doi.org/10.1016/j.appet.2013.01.020>
- ¹³ Allcott, H. (2015). Site Selection Bias in Program Evaluation. *The Quarterly Journal of Economics*, 130(3), 1117–1165. <https://doi.org/10.1093/qje/qjv015>
- ¹⁴ Pritchett, L. (2002). It pays to be ignorant: A simple political economy of rigorous program evaluation. *The Journal of Policy Reform*, 5(4), 251–269. <https://doi.org/10.1080/1384128032000096832>
- ¹⁵ Hummel, D., & Maedche, A. (2019). How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. *Journal of Behavioral and Experimental Economics*, 80, 47–58. <https://doi.org/10.1016/j.socec.2019.03.005>.
- ¹⁶ Allcott, H., & Kessler, J. B. (2019). The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons. *American Economic Journal: Applied Economics*, 11(1), 236–276. <https://doi.org/10.1257/app.20170328>
- ¹⁷ Ozaki, T., & Nakayachi, K. (2020). When Descriptive Norms Backfire: Attitudes Induce Undesirable Consequences during Disaster Preparation. *Analyses of Social Issues and Public Policy*, 20(1), 90–117. <https://doi.org/10.1111/asap.12195>
- ¹⁸ Costa, D. L., & Kahn, M. E. (2013). Energy Conservation “Nudges” and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment. *Journal of the European Economic Association*, 11(3), 680–702. <https://doi.org/10.1111/jeea.12011>
- ¹⁹ Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627–668. <https://doi.org/10.1037/0033-2909.125.6.627>
- ²⁰ See Pink, D. H. (2011). *Drive: The Surprising Truth About What Motivates Us*. Penguin; and Kohn, A. (1999). *Punished by Rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes*. Houghton Mifflin Harcourt.
- ²¹ Goswami, I., & Urminsky, O. (2017). The dynamic effect of incentives on postreward task engagement. *Journal of Experimental Psychology: General*, 146(1), 1–19. <https://doi.org/10.1037/xge0000206>
- ²² Angrist, J., Lang, D., & Oreopoulos, P. (2009). Incentives and Services for College Achievement: Evidence from a Randomized Trial. *American Economic Journal: Applied Economics*, 1(1), 136–163. <https://doi.org/10.1257/app.1.1.136>
- ²³ Goswami, I., & Urminsky, O. (2017). The dynamic effect of incentives on postreward task engagement. *Journal of Experimental Psychology: General*, 146(1), 1–19. <https://doi.org/10.1037/xge0000206>
- ²⁴ Urminsky, O. & Goswami, I. (2016) One Size Does Not Fit All: Importance of Donor Type on setting Default Amounts for Charitable Donations. *University of Chicago. Working paper*.
- ²⁵ Sunstein, C. R. (2017, December 28). Don't Underrate the Power of the Default Option. *Bloomberg.Com*. <https://www.bloomberg.com/opinion/articles/2017-12-28/don-t-underrate-the-power-of-the-default-option>
- ²⁶ Hummel, D., & Maedche, A. (2019). How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. *Journal of Behavioral and Experimental Economics*, 80, 47–58. <https://doi.org/10.1016/j.socec.2019.03.005>

-
- ²⁷ Johnson, E. J., & Goldstein, D. (2003). Do Defaults Save Lives? *Science*, 302(5649), 1338–1339. <https://doi.org/10.1126/science.1091721>
- ²⁸ Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2009). The importance of default options for retirement saving outcomes: Evidence from the United States. In *Social security policy in a changing environment* (pp. 167-195). University of Chicago Press.
- ²⁹ Defaults Are Not the Same by Default. (2019, April 16). *Behavioral Scientist*. <https://behavioralscientist.org/defaults-are-not-the-same-by-default/>
- ³⁰ Goswami, I., & Urminsky, O. (2016). When should the Ask be a Nudge? The Effect of Default Amounts on Charitable Donations. *Journal of Marketing Research*, 53(5), 829–846. <https://doi.org/10.1509/jmr.15.0001>
- ³¹ Urminsky, O. & Goswami, I. (2016) One Size Does Not Fit All: Importance of Donor Type on setting Default Amounts for Charitable Donations. *University of Chicago. Working paper*.
- ³² Goswami, I., & Urminsky, O. (2020). No Substitute for the Real Thing: The Importance of In-Context Field Experiments in Fundraising. *Marketing Science*, 39(6), 1052–1070. <https://doi.org/10.1287/mksc.2020.1252>
- ³³ Leonhardt, D. (2008). What Makes People Give. *The New York Times*, 9.
- ³⁴ Andreoni, J. (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *The Economic Journal*, 100(401), 464–477. <https://doi.org/10.2307/2234133>
- ³⁵ Karlan, D., & List, J. A. (2007). Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment. *American Economic Review*, 97(5), 1774–1793. <https://doi.org/10.1257/aer.97.5.1774>