



Changing Behavior Beyond the Here and Now

Faculty Research Working Paper Series

Todd Rogers

Harvard Kennedy School

Erin Frey

Harvard Business School

February 2014
RWP14-014

Visit the **HKS Faculty Research Working Paper Series** at:
<https://research.hks.harvard.edu/publications/workingpapers/Index.aspx>

The views expressed in the **HKS Faculty Research Working Paper Series** are those of the author(s) and do not necessarily reflect those of the John F. Kennedy School of Government or of Harvard University. Faculty Research Working Papers have not undergone formal review and approval. Such papers are included in this series to elicit feedback and to encourage debate on important public policy challenges. Copyright belongs to the author(s). Papers may be downloaded for personal use only.

Changing Behavior Beyond the Here and Now

Todd Rogers¹ and Erin Frey²

¹Harvard Kennedy School

²Harvard Business School

Published in: Blackwell Handbook of Judgment and Decision Making (2015)

Words: 9,075

Acknowledgements: We thank Hunt Allcott, Sendhil Mullainathan, Max Bazerman, Francesca Gino, Hunter Gehlbach, Mike Norton, Betsy Paluck, Richard Zeckhauser, and David Halpern and the Behavioral Insights Team in the United Kingdom for helpful feedback.

ABSTRACT:

We develop a framework for thinking about and influencing how behavioral science-informed interventions operate through time. Intervention-behavior lags occur when interventions are administered at one point in time, but the behaviors they target are performed in the future. Marginal benefits to continued treatment occur when successive rounds of a repeated intervention each impact the target behavior. Persistence occurs when treatment effects persist even after interventions have been discontinued. We discuss the features of interventions that influence intervention-behavior lags and the marginal benefits to continued treatment, and propose several pathways through which interventions may generate persistent treatment effects.

Keywords: Nudges; Persistence; Field Experiments; Interventions; Time; Behavior Change

Behavioral science insights are increasingly being used to enhance interventions that improve societal well-being. These insights are decreasing smoking, obesity, energy usage, littering, and conflict; and they are increasing savings, voting, medication adherence, donation, educational achievement, and tax compliance. As the applications of this science have developed, a class of questions has emerged related to how time affects the impact of interventions. This manuscript explores several of these questions. The discussion has practical implications for policy-makers aiming to influence behavior, while also laying out a framework for future research and theory about behavioral science-informed interventions.

Over the past sixty years, behavioral decision researchers have uncovered underlying patterns in how and when people deviate from the principles of rationality, and have developed a science of judgment and decision-making (e.g., Kahneman & Tversky, 1979; Gilovich, Griffin, and Kahneman, 2002). In recent years this work has taken on a decidedly prescriptive thrust (Thaler & Sunstein, 2003, 2009; Camerer, Loewenstein and Prelec 2003), naturally suggesting innovations in how interventions are structured. In parallel – and in part because of the increasing application of this work – there has been an explosion in the use of randomized field experiments across the social sciences (e.g., Schultz et al., 2007; Madrian & Shea, 2001; List, 2011; Chattopadhyay & Duflo, 2004). The current manuscript combines these two developments: we use the insights and perspective of behavioral science to analyze how field interventions to improve societal well-being work over time. We will focus on the following three aspects of time and interventions.

- I. **Intervention-Behavior Lags:** Interventions are sometimes administered temporally close to when the target behavior is to be performed, and sometimes they are administered temporally far from when the target behavior is to be performed. Consider, for example, interventions to make drivers fasten their seatbelts. The audible “beep” reminder that an unbuckled driver hears upon turning on the car is an example of an intervention with a short intervention-behavior lag. The graphic advertisements depicting car crashes that a driver sees hours, days, or weeks before they drive is an example of an intervention with a longer intervention-behavior lag. In this manuscript, we will discuss features of interventions that affect their success at bridging this intervention-behavior lag.
- II. **Marginal Benefits to Continued Treatment:** Interventions are often administered repeatedly over time to maintain behavior change. For example, some utilities give customers feedback about their energy use as compared to their neighbor’s energy use every month. Intuitively, it is not difficult to imagine why the umpteenth application of an intervention would not be as effective as the first. This manuscript will explore why interventions might or might not continue to have an impact with each successive round of treatment.
- III. **Persistence:** Interventions are sometimes administered during a finite window but with the objective of inducing long-term, persistent change, even after the treatment itself has ended. For example, some incentive-based interventions aim to induce individuals to continue a regular exercise regimen, even after the incentives have been removed. This manuscript will discuss several pathways that might lead to the persistence of treatment effects after an intervention has been discontinued.

In this manuscript we discuss interventions targeting a range of societal challenges. Some of these challenges arise when individuals intend to engage in virtuous or healthful behaviors, but then fail to follow-through on these intentions (e.g., saving money or exercising). Other challenges occur when people hold mistaken beliefs that affect their behaviors, and these behaviors reduce social welfare (e.g., mistakenly believing that one is the only person on campus who tends to not binge drink and therefore

engaging in more binge drinking than one would otherwise prefer). Yet other challenges arise when people's behaviors impose negative costs on society but create personal benefits (e.g., commons problems). It is beyond the scope of this manuscript to discuss the ethics of behavioral interventions (for an accessible overview see Sunstein, 2013). Given that interventions for social value creation are developed and are administered, our aim is to understand how advances in behavioral science can help us understand the relationships between interventions, time, and treatment effects.

Societal challenges arise for many reasons. Throughout this manuscript we will refer to four of the many factors that may contribute to the existence of these challenges: inattention/forgetting, time inconsistent preferences, mistaken beliefs, and external obstacles. These are not the only factors that play a role in societal challenges, but we select these four to illustrate our arguments. First, people sometimes forget to perform behaviors. This may be caused by a failure of memory (Ellis, 1996; Schacter, 1999), by a lack of attention at the appropriate moment (McDaniel & Einstein, 2000) or by excessive cognitive busyness at the time the behavior is to be performed (Sweller, 1988). Second, people may exhibit inconsistent preferences over time. Not only do people have evolving preferences, they also have systematically different preferences across time (Laibson, 1997; Milkman, Rogers, Bazerman, 2008). That is, people's preferences towards future behaviors tend to be relatively biased towards what people believe they virtuously "should" do, whereas their preferences for current behaviors tend to be relatively biased towards what people indulgently "want" to do – and this can be exacerbated when people's attention and self-control are taxed (Shah, Mullainathan and Shafir, 2012; Muraven, Tice, and Baumeister, 1998). Third, people sometimes have mistaken beliefs that can interfere with welfare-improving behaviors. These could result from motivated reasoning (Kunda, 1990), biased perception (Lord, Ross, and Lepper, 1979; Prentice and Miller, 1993), or a simple lack of knowledge. Finally, external obstacles may prevent people from following through on intended behaviors. These could be physical (i.e., a broken car that prevents people from driving to their polling place to vote), temporal (not having enough time to perform the behavior), or social (i.e., significant others monitoring and expressing disapproval about the behavior). Inattention/forgetting, time inconsistent preferences, mistaken beliefs, and external obstacles all contribute to individuals failing to perform welfare-improving behaviors, thereby giving rise to the need for interventions informed by behavioral science.

Two clarifications before we proceed. First, we should define what we mean by a "behavioral intervention." We have a broad conception of what qualifies: an intervention with features that reflect behavioral science insights aimed at actively changing targets' cognitive, social, and informational decision context in order to influence targets' behaviors. Unlike the mainstream economic approach, these interventions generally involve more than direct financial incentives (for a more extensive discussion of the difference between a behavioral approach and an economic approach to interventions, see Amir et al., 2005). Second, we note that while we will discuss a range of field research from the behavioral sciences that report short-term and long-term effects, we do not aspire to offer an exhaustive review of such work. Though worthwhile, such an endeavor would be beyond our specific objective: to describe the dynamics of behavioral intervention treatment effects over time, and to unpack some of the psychological features and processes underlying these dynamics.

I. Intervention-Behavior Lag

Behavioral interventions vary widely in the degree to which there is a lag between the moment that they are administered and the moment that the target behavior is to be performed. Consider several objectives those who develop interventions often have: they aim to induce diners to properly dispose of their trash at the end of their meals; they aim to encourage people to vote in an election two weeks in the future; they aim to induce people to save for their retirements each pay cycle for the next ten years. Below we describe features that are likely to make interventions serving these types of objectives more effective at inducing target behaviors soon after the interventions are administered (short intervention-behavior lag);

we then examine the features of behavioral interventions that are likely to make them more effective at bridging time and inducing target behaviors performed temporally distant from when the interventions are administered (long intervention-behavior lag).

Features likely to induce behavior soon after intervention is administered

Often practitioners and policy makers design behavioral interventions with short intervention-behavior lags—the intervention is administered at the exact moment that the target behavior is performed. Moreover, the rapid expansion of mobile technology platforms is making this type of intervention increasingly possible.

Below we describe three features of behavioral interventions that are especially well suited to induce people to perform target behaviors at the exact moments when the interventions are administered. Because these features work primarily by reducing the likelihood of inattention/forgetting in the moment when target behaviors are to be performed, these features are not likely to be particularly useful at inducing target behaviors that must be performed long after the treatment is administered.

Interventions that default individuals into a behavior. Interventions that leverage defaults automatically enroll individuals into a program or course of action; if individuals do not wish to participate, they must proactively undertake specified steps so as to be removed from the program or course of action. This intervention feature occurs at exactly the moment when the target behavior is to be performed. In some sense, defaults preempt the need to induce a target behavior altogether.¹ This particular feature is hard to imagine as part of a behavioral intervention administered temporally distant from when the target behavior is to be performed.

Interventions leveraging defaults have been implemented widely by policy makers (e.g., Johnson & Goldstein, 2003), and the domain that has experienced some of the most powerful effects from this feature involves default-enrolling new employees into 401(k) savings plans (Madrian & Shea, 2001). This savings plan involves directing pre-tax income to an investment account that accumulates savings, tax-protected, until the saver reaches a certain age. 401(k) accounts are usually seen as good savings vehicles because they allow assets to grow pre-tax, and savers often receive additional matching contributions from their employers. Despite these advantages millions of eligible workers do not participate in their 401(k) savings plans. Behavioral researchers have found that defaulting new employees into these accounts when they complete the required new employee paperwork (e.g., tax information, whether one wants direct deposit, which healthcare plan one will enroll in, etc.) dramatically increases enrollment. In one study, for example, auto-enrollment increased participation from 37% to 86% (Madrian & Shea, 2001). The behavioral intervention (default enrollment) occurs at the exact moment that employees are completing paystub-related paperwork. All of the information needed to participate in the 401(k) is already on other forms being filled out; therefore, signing up for a 401(k) plan requires no additional time or effort. By embedding the intervention in the moment of choice, individuals will be induced to perform the target behavior immediately after they are exposed to the intervention, thereby leveraging a very short intervention-behavior lag².

¹ In these cases, defaults may be effective interventions not only for challenges caused by inattention/forgetting, but also for challenges arising from time inconsistent preferences. In a sense, defaults could be seen as preempting the need for individuals to perform target behaviors in the future, which is especially consequential when those behaviors involve time inconsistent preferences.

² It should be noted that defaults may also prevent future attention from being directed toward the behavior. This may have consequences that are not welfare enhancing if the individuals are defaulted into a sub-optimal

Interventions that momentarily direct bounded attention. People have limited attentional capacity, and therefore cannot attend to all possible stimuli in any given moment (Simons & Chabris, 1999; Chugh & Bazerman, 2007). Individuals are also more likely to be influenced by what they are paying attention to; thus, if a behavioral intervention can direct individuals' attention to a particular target behavior, the individuals will be more likely to perform that behavior.

The timing of behavioral interventions designed to capture and direct attention is significant. Nickerson (2007) found that individuals who received a reminder to vote (via telephone) in the last few days before an election were more likely to vote than those who received the same phone reminder a week before the election. Sometimes even minutes can influence the effectiveness of a reminder. Austin, Sigurdsson and Rubin (2006) found that reminding drivers to use their seat belts immediately before they drove away significantly increased seat belt compliance. However, when the reminders were administered five minutes before drivers drove away these reminders had no effect on seat belt usage. Thus, momentarily directing bounded attention can induce individuals to perform a target behavior immediately after a behavioral intervention.

Interventions that momentarily change accessible thoughts. Behavioral interventions can also induce a target behavior immediately by temporarily altering what thoughts are accessible to an individual. We define thoughts to include perceptions, beliefs, decision rules or goals. Like defaults and directing bounded attention, behavioral interventions that involve changing accessible thoughts are most effective when the lag between intervention and target behavior is short. Here we discuss two of the ways that thoughts can be momentarily changed so as to immediately induce target behaviors: the framing of risky choices, and changing the accessibility of target behavior relevant goals (i.e., cognitive "priming").

In situations of uncertainty, an individual's decisions can be influenced by the way that gains and losses are presented. Tversky and Kahneman (1981) demonstrated that individuals tend to be risk-seeking when presented with a loss framing and risk-averse when presented with a gain framing—even when those two frames convey the same objective information. This type of framing affects more than an individual's judgment; loss and gain framing has also been used to influence an individual's behaviors. For example, Detweiler et al. (1999) found that when they presented beachgoers with information about sunscreen use and skin cancer in a loss-frame ("not using sunscreen increases your risk for skin cancer and prematurely aged skin"), beachgoers were more likely to purchase sunscreen that day than those who were presented with the same information in a gain-frame ("using sunscreen decreases your risk for skin cancer and prematurely aged skin"). This illustrates how the framing of a choice can alter people's thoughts (risk attitudes) in a specific moment³, and can thereby change their immediate behavior.

contribution level, or if individuals' preferences change over time; for example, if individuals are defaulted into being organ donors, but then over time they decide that they no longer want to be donors, they may never actually change their donor status because their attention will not be directed toward making another active choice about whether to be a donor.

³ With this example, we are not suggesting that loss-framing cannot enduringly change perceptions of risk (i.e., Ganzach & Karsahi, 1995). There has been some research showing longer-term effects from loss-framing interventions, though how these longer-term effects arise are not well unpacked in that research. We discuss possible pathways for longer-term effects later in this manuscript.

Making specific thoughts cognitively accessible during the exact moment when a behavior is to be performed can also influence the impact of a behavioral intervention.⁴ For example, priming people with food-related constructs by showing them commercials for snacks as opposed to showing them commercials for non-food items can induce them to eat more in that moment (Harris, Bargh, and Brownell, 2009). Similarly, LeBoeuf, Shafir, and Bayuk (2010) found that people's preferences can be influenced by the self-identities that are cognitively accessible at the moment of choice. In one experiment, participants primed with one self-identity expressed preferences that were congruent with that identity; however, when a different self-identity was primed immediately after, the participants expressed preferences that were congruent with the newly primed self-identity—even when the second set of preferences contradicted the first. Thus, making specific thoughts salient in the moment when a target behavior is to be performed can trigger the target behavior.

Changing thoughts can contribute to the effectiveness of a given behavioral intervention, but only for as long as those changes endure. If a target behavior is to occur temporally near to when a behavioral intervention is administered, a fleeting change in what is cognitively accessible may influence the behavior. However, if a target behavior is to occur temporally far from when a behavioral intervention is administered, a fleeting change in an individual's thoughts is not likely to influence the target behavior. Of course, some thoughts can be enduringly changed—we will discuss that later in the manuscript—but here we emphasize the impact of fleeting changes in the accessibility of these thoughts.

Defaulting individuals into a choice, capturing and directing individuals' attention, and momentarily changing what thoughts are cognitively accessible to individuals in a given moment are especially likely to affect behavior when they are associated with short intervention-behavior lags. These features of interventions can help induce a target behavior immediately after a behavioral intervention is applied because they reduce the likelihood that individuals will fail to attend to the behavior (or will forget to perform the behavior). As such, interventions with these features are likely to be especially effective at addressing challenges caused by inattention and forgetting.

Features likely to bridge time

Although many practitioners would prefer to administer behavioral interventions immediately before the target behavior is to be performed, often this is not possible. In fact, many of the most important, expensive, and common behavioral interventions need to “bridge time.” That is to say that the interventions must be administered in moments temporally distant from the moment when individuals are likely to perform the target behaviors.

Features that help interventions bridge the temporal gap between intervention and behavior are especially likely to be effective at addressing challenges caused by inattention and forgetting, and those caused by preferences for the present and preferences for the future differing (i.e., time inconsistent preferences). Intervention features that bridge time may help individuals re-direct their attention or activate relevant memory contents at appropriate times, thereby addressing inattention and/or forgetting challenges. Similarly, intervention features that bridge time may increase the likelihood that people's preferences for their future behaviors influence their actual future behaviors, thereby mitigating the impact of time inconsistent preferences.

⁴ This is consistent with query theory's central premise that preferences are shaped by the order in which information is queried in the mind; specifically, the first set of information people pull to attention tends to dominate preferences because this information tends to become especially accessible (Johnson, Haubl and Keinan, 2007).

Three features of behavioral interventions allow interventions to bridge time and affect behaviors that are temporally distant from the moment when the intervention is administered.

Interventions that chronically direct bounded attention. Behavioral interventions that increase the frequency with which individuals think about the target behavior can address challenges related to inattention and forgetfulness. As described above, if people fail to perform a behavior because of inattention/forgetting, then inducing people to think about a target behavior at the exact moment that it should be performed will increase the likelihood that people will, in fact, perform it. When discussing short intervention-behavior lags, we noted that momentarily increasing cognitive accessibility of a target behavior can induce the behavior. We also noted that interventions that momentarily direct bounded attention to the target behavior are likely to produce short intervention-behavior lags. However, these fleeting changes in attention generally do not endure. In order for a behavioral intervention to bridge time, the behavior must chronically capture individuals' attentions.

Social accountability can be leveraged to make individuals chronically attend to target behaviors. To illustrate this feature, consider a recent voter mobilization field experiment by Gerber, Green, and Larimer (2008). In one experimental condition, households were sent mailers that displayed the voting history of those who resided in the household, and also the voting history of their neighbors. The mailing also reported that one's neighbors had received a similar mailing, and that the recipient and his or her neighbors would receive a follow-up letter after the election that would show who in the neighborhood had voted, and who had failed to vote. This behavioral intervention aimed to create the feeling of social pressure: if the target individuals failed to follow through on the target behavior of voting, their neighbors would know and would sit in quiet judgment. This get-out-the-vote mailing was nearly five times more effective than a standard voter encouragement mailing tested in the same experiment. In a related field experiment, Rogers and Ternovski (2014) found that a mailer leveraging social accountability increased the effectiveness of a psychologically sophisticated mailer by more than 50%, and those who reported caring most about civic life were by far the most influenced by this intervention. In both of these examples, the threat of social accountability may have made the target individuals more likely to chronically think about the target behavior (to vote),⁵ and therefore become more likely to actually perform that behavior.

Interventions that link a performance environment with an intention. Behavioral interventions that induce people to cognitively link their intention to perform a target behavior to the environment in which the behavior is to be performed ("performance environment") may be especially potent at bridging time when the underlying reason a behavior is not being performed is inattention/forgetfulness. Whereas the feature described above, chronic cognitive accessibility, entails inducing targets to attend to a behavior at greater frequency over time, this feature uses the performance environment to make the intention cognitively accessible at exactly the right moment. Plan-making is an example of a behavioral intervention that incorporates this feature. Not only does it encourage individuals to develop strategies for overcoming

⁵ These studies did not directly measure the cognitive state of increased accessibility, nor did they show that such a state mediates the observed treatment effects. Such an interpretation is plausible, though, and these studies are cited to bring to life how this intervention feature could bridge time. For example, recent evidence suggests that individuals experience prolonged emotions as a result of these social accountability interventions: such treatments have generated public outrage many days after they were administered by political organizations (Mataconis, 2012).

possible logistical obstacles, but on a cognitive level, plan-making also helps people to remember their intentions at appropriate times (Gollwitzer, 1999; Gollwitzer & Sheeran, 2006).

The formation of implementation intentions through plan-making is one type of intervention that may bridge time by linking the target behavior to the performance environment. For example, Nickerson and Rogers (2010) conducted a get-out-the-vote experiment where individuals in one condition were called and asked several questions about the details of how they planned to vote on election day: when they would vote, how they would get to their polling place, and where would they be coming from when they went to vote. This plan-making intervention was more than twice as effective as a standard get-out-the-vote call. In essence, this intervention created a link between the individual's intention (to vote) and the performance environment (the time and place they would be on Election Day when they were to head to their polling place), thereby allowing this intervention to bridge time. Plan-making has also been used to induce individuals to exercise (Milne, Orbell, & Sheeran, 2002), get a flu shot (Milkman, Beshears, Choi, Laibson, and Madrian, 2011), diet (Achtziger, Gollwitzer, & Sheeran, 2008), and get a colonoscopy (Milkman et al., 2012).

Interventions that allow individuals to pre-commit to a future behavior. Behavioral interventions that target behaviors that involve time-inconsistent preferences are more likely to bridge time if they induce people to adopt a commitment device. Commitment devices are voluntary arrangements that individuals enter into to help them follow through on a future behavior that may require self-control (Bryan, Karlan, and Nelson, 2010). In these arrangements, individuals voluntarily agree to incur a cost or penalty if they do not perform the specified target behaviors by a pre-specified future date. For example, leaving one's credit cards at home and bringing a limited amount of cash when going out for the evening serves as a commitment device designed to limit spending. Acquiring additional money—by borrowing from a friend, for example—would incur psychological and possibly financial costs. This intervention bridges time, as the intervention (choosing to leave credit cards at home and bringing only cash) likely affects a target future behavior (spending during the evening).⁶

The strength, design, and longevity of commitment devices can vary widely. “Immutable” commitment devices are those that make individuals incur unavoidable costs if the target behavior is not performed. For example, some alcoholics take metabolism-inhibiting drugs (such as Antabuse) that trigger severe nausea and vomiting if they drink alcohol in the near future. Such an unpleasant physical state may be considered a large cost for failing to perform the target behavior (avoiding alcohol consumption). “Mutable” commitment devices, on the other hand, cause individuals to incur costs and inconveniences if the target behaviors are not performed. For example, alcoholics may choose to not keep alcohol in their homes; if they fail to perform the target behavior (avoiding alcohol consumption), they incur the relatively light cost of having to make a trip out to purchase alcohol. It is likely that immutable commitment devices are particularly effective for interventions that target behaviors driven by time-inconsistent preferences, while mutable commitment devices may be particularly effective for interventions that address inattention/forgetfulness. If people's preferences change over time, then a minor cost is unlikely to induce them to follow their initial preferences; however, a minor cost may be enough to direct people's attention to a target behavior they might have been inattentive about. For example, Schwartz et al. (2014) found that offering people an immutable commitment device (forfeiting a

⁶ It should be noted that the decision calculus people engage in when commitment devices are employed often leverage other biases. For example, some commitment devices change people's decision calculus by leveraging sunk costs (Della Vigna & Malmendier 2006).

sizable rebate if they did not purchase enough healthy food) substantially increased the purchase of healthy foods. Regardless of the structure, commitment devices contribute to bridging time by increasing the costs of failing to perform target behaviors at pre-specified future moments.

Interventions that enduringly change consequential thoughts. A final feature that likely helps behavioral interventions bridge time is if they enduringly change the thoughts that influence how individuals behave. While above we discussed that the accessibility of thoughts (perceptions, beliefs, decision rules and goals) can be fleetingly changed, here we discuss that sometimes the content of thoughts can be enduringly changed. If specific thoughts are consequential to a behavior, then changing those thoughts may lead to enduring changes in behavior. These behavior changes can result from correcting inaccurate thoughts, or from introducing new thoughts. For that reason, interventions involving this feature are likely to be particularly effective for addressing challenges affected by mistaken beliefs.

In some cases, simply correcting inaccurate but important beliefs can change mental content in ways that help interventions bridge time. For example, surveys show that a surprising fraction of American voters believe that who they vote for in elections is public record. Of course, this is a factually incorrect belief, as the United States has used secret ballots in federal elections since 1892. Recent work has shown that explaining to people who consistently do not vote that their ballots are, in fact, secret powerfully increases their likelihoods of casting a ballot (Gerber et al., 2013). This work suggests that one of the restraining forces on voter participation is the false belief that ballots are not secret. This is the kind of erroneous mental content that, once changed, could produce behavior changes after long intervention-behavior lags.

Features of interventions that can cause immediate behavior change differ from those that can bridge time. Here we have highlighted three features of behavioral interventions that are likely to induce a target behavior immediately (have a short intervention-behavior lag) and four features that are likely to help interventions bridge time (have a long intervention-behavior lag). Our goals were two fold. Firstly, to highlight the importance of the temporal relationship between when an intervention is administered and when the target behavior is to be performed. And secondly, to describe some key features of interventions that might be particularly important depending on the relationship between the moment when the intervention is to be administered and when the target behavior is to be performed. Of course, there are likely other features we could have discussed, and still others that are as yet not known and future research will uncover.

II. Marginal Benefit to Continued Treatment

In this section we turn our attention to the marginal benefit to continued application of a behavioral intervention. Specifically, we are interested in how additional administration of an intervention could result in behavior change above and beyond the behavior change that might persist from whatever treatment had already been administered. The absence of marginal benefits to continued treatment would mean that each additional round of a behavioral intervention would have no positive marginal impact relative to not administering it.

From a cognitive perspective, it is unlikely that repeated instances of a single intervention would continue to have the same effect each time and produce the same level of a target behavior. Research in classical conditioning demonstrates that humans (and other organisms) quickly become desensitized to repeated exposure to a given stimuli; in general, an individual's behavioral response to the first instance of a stimuli tends to be more pronounced than his/her behavioral response to the n th repetition of that same stimuli (Rankin et al., 2009; Thompson & Spencer, 1966). Yet, despite the fact that humans habituate to

repeated exposure to the same stimuli, researchers have at times observed positive marginal benefits to continued treatment— administering additional instances of the same behavioral intervention has produced more of the target behavior relative to not administering the additional instance.

Below we identify three features of behavioral interventions that may generate marginal benefits to continued treatment.

Interventions that resist attention habituation. Research on habituation finds that people usually become desensitized after repeated exposure to a stimulus over time. However, there are four qualities of a stimulus that prevent desensitization and habituation. First, when the intervals between repetitions are long, individuals do not habituate to a stimulus as much; when the repetition does occur, the stimulus is experienced as sufficiently novel so as to immediately direct an individual's attention (Rankin et al., 2009). Second, repeated exposure to the same stimulus may not cause habituation if the sensory properties of the stimulus are somewhat dynamic (McSweeney, 2004).⁷ Third, individuals tend to not habituate to repeated exposure to a stimulus if the stimulus is sufficiently intense (Rankin et al., 2009). Lastly, humans resist habituation to repeated exposure to a stimulus if it is presented at unpredictable intervals (McSweeney, 2004).

Interventions that produce incomplete or temporary change in behavior. A different way of thinking about why continued application of a behavioral intervention may show a marginal benefit is to look at the extreme case of when continued treatment would show no marginal benefit. This could occur if the first application of the intervention completely and enduringly changed the focal behavior, rendering continued treatment redundant. For example, consider the secret ballot study by Gerber et al. (2011), described above. In that study, individuals incorrectly believed that who they voted for in an election was a matter of public record. When these individuals were disabused of that false belief and came to understand that their ballots were, in fact, secret, they became significantly more likely to vote. Repeating this intervention would likely have had trivial marginal impact on voting because targets' beliefs had already been corrected with the first application of the intervention. Conversely, interventions that induce incomplete changes in behavior are more likely to exhibit marginal benefits to continued treatment.

Interventions that induce minimal resistance or distrust. In some instances, individuals may become suspicious of the source, content, and intent of the intervention (McGuire, 1964; Friestad & Wright, 1994), particularly if the intervention is novel, unusual, or otherwise attracts unusual scrutiny by the individuals. If individuals become suspicious, they may choose not to perform the target behavior. While this could be of concern for a one-off treatment, repeated administration of a treatment may render it especially vulnerable to this greater skepticism as it may prompt heightened reflection on the source's motives. This heightened skepticism and distrust might result in continued application of a behavioral

⁷ For an illustration of how a dynamic intervention that is continually administered could be of positive marginal benefit consider a recent study on how engrossing audiobooks have been used to induce exercise. Milkman, Minson, & Volpp (2012) made these audiobooks available to participants only when participants were at the gym. They found that bundling pleasant and addictive experiences (e.g., listening to the audiobook) with a virtuous but aversive experience (e.g., exercising) maintained increased gym attendance over eight weeks. Though they did not assess how much the exercise behavior would have persisted if they had discontinued making the audiobooks available, one can speculate that the intervention's effects would have rapidly decayed. In fact, they found that their effect on exercise was fragile: once participants returned from a long break (e.g., the Thanksgiving holiday) the effect no longer survived.

intervention not generating marginal benefits. Interventions that depend on peripheral routes to behavior change, as opposed to central routes to behavior change (Petty & Cacioppo, 1986; Petty, 2013), might be especially vulnerable to this resistance. Interventions that work through central routes involve persuasion that is based on the quality of arguments or message content; interventions that work through peripheral routes influence attitudes and behaviors through non-message relevant cues, such as the perceived expertise of the source, the listener's mood, and contextual factors. Interventions relying on the central route are likely to be transparent about their purposes. Therefore individuals likely recognize from the beginning that these treatments intend to change their behaviors and so repeated treatment will likely not lead individuals to second guess the objectives of the interventions – the objectives are explicit and transparent. On the other hand, interventions relying on peripheral routes are less likely to be transparent about their purposes. This means that repeated administration of a treatment may induce individuals to elaborate on (and view with greater skepticism) the purposes of the interventions. Growing aware of this indirectness may induce elaboration and reactance against the objectives of these interventions (Petty and Cacioppo, 1977).

III. Persistence

In this section we take a different perspective on how time affects the impact of behavioral interventions. Rather than examining features of interventions, here we use a behavioral science perspective to analyze how treatment effects from any intervention can persist after the intervention has been discontinued. That is, what are the psychological and structural pathways through which behaviors can enduringly change as a result of interventions? It is notable that the logic of how a behavior change might persist over time is often unrelated to the logic of how the intervention caused the initial behavior to change. For the remainder of this discussion “persistence” refers to the repeated performance of a target behavior without the administration of additional intervention, and usually after the first performance of that behavior.

Despite the importance of understanding how and when interventions will yield persistent behavior change, most field research assesses outcomes just once and often shortly after a treatment has been administered. While this research design can be expedient, relatively cost effective, and address many important research questions, it fails to shed light on the temporal dynamics and longevity of the treatment effects these interventions bring about. There are some exceptions, though. Research investigating social-belonging and mindset interventions in education (Walton & Cohen, 2011; Dweck, 2007) and social comparison interventions in environmental conservation and energy use (Allcott & Rogers, in press; Ferraro, Miranda, and Price, 2011) have examined and discovered persistent treatment effects, even several years after interventions have been discontinued. In the domain of health, some interventions also show evidence of persistence. Charness and Gneezy (2009) found that people who received monetary incentives to exercise continued to go to the gym for weeks after the incentives ended.

However, many other interventions have failed to generate lasting treatment effects. For example, John et al. (2011) found that individuals lost a significant amount of weight while they participated in an eight-month weight-loss program involving lotteries. Once the program ended, however, those individuals regained all of the lost weight over the next four months. In this case, the effects of the intervention (the weight loss program) were not persistent, as the target behavior (weight loss or weight maintenance) ceased when the intervention was discontinued. Incentive-based smoking cessation programs show similar results. A meta-analysis of seventeen studies conducted by Cahill and Perera (2008) concluded that “there is no compelling evidence...that competitions or incentives improve long-term smoking cessation...Several studies identified higher early and medium-term quit rates for the intervention groups,

but these encouraging signs generally did not survive into long-term abstinence.” On the other hand, Volpp et al. (2009) found that their incentive program both increased smoking cessation six months after program enrollment, and also generated persistence (i.e., smoking abstinence six months after the cessation program ended), but this finding is an outlier among smoking cessation research.

This inconsistent pattern begs the question: why do some interventions generate persistent effects, while others do not? An array of factors may contribute to the persistence of treatment effects from interventions. In this section, we identify five pathways⁸ through which interventions might induce persistent behavior change. These pathways fall into two categories: internal-to-self (habit and changes to mental contents) and external-to-self (social reinforcement, changes to future costs, and rip currents). It is likely that the persistence of treatment effects from interventions is caused by the simultaneous influence of several of these pathways. That said, each pathway is conceptually distinct, and so merits individual description.

Habit pathway. One internal-to-self way that an intervention could induce persistent behavior change is through the development of psychological habits. If an intervention causes people to repeatedly perform the same target behavior in the same environment, the individuals could develop a memory-based association between the behavior and the environment in which the behavior is performed. This can create associations such that when the individuals are in the particular performance environment, they automatically perform the target behavior (Ouellette & Wood, 1998). These associations can cause the performance environment to automatically induce the performance of the target behavior without conscious awareness, or it may induce conscious recall of the need to perform the target behavior. Both routes could lead to persistent behavior change.

Habit may explain the persistent effects observed from some energy-related interventions. For example, imagine a person puts a note “turn off lights!” on the door most used by others in her family to exit the house. The first few times the family members leave the house the note directs their attention to turning off the lights, and they do. Eventually, after repeating this behavior many times at this exact moment, exiting through that door may come to be associated with turning off the lights. Even without noticing the note, the family members may automatically turn off the lights. This psychological habit could persist after the sign is no longer noticed, and even after it is no longer in place, supporting a persistent change in behavior. Psychological habits have been cited as contributors to the repetition of high frequency health-related behaviors, such as diet (Rothman, Sheeran, & Wood 2009) and exercise behaviors (Charness & Gneezy, 2009; Acland & Levy 2010; Wood, Tam, & Witt 2005).

Changing mental contents pathway. In the section on intention-behavior lags we discussed how enduringly changing thoughts can help behavioral interventions bridge time. Not only can changes in thoughts, beliefs, attitudes, or interpretations help interventions bridge time, but such changes can also increase the likelihood that target behaviors will continue to be performed in the future, long after the interventions have been discontinued. Behaviors may persist over time if interventions permanently change beliefs or interpretations that are causally consequential to the behaviors. This may arise by

⁸ In the previous sections we discussed *features of* behavioral interventions that influence intervention-behavior lags and marginal benefits to continued treatment. In this section, we identify *pathways through which* treatment effects can persist over time. We are interested in teasing out why behavior change might persist after any kind of intervention is discontinued, and not necessarily the features of an intervention that might lead to persistence. Whereas the first two sections focused on specific aspects of interventions, this section discusses the processes, external to interventions themselves, through which persistence might arise.

replacing existing beliefs with different beliefs, or by creating beliefs where they did not previously exist. Persistent behavior change may also occur if interventions change how people interpret ambiguous information (Ross & Nisbett, 1991). Like habit formation, both changes to beliefs and changes in interpretations are pathways that are internal-to-self. Below we first discuss how replacing *beliefs* may lead to persistent behavior change; we then describe how changing the way ambiguous information is *interpreted* can also generate persistence.

Beliefs: One particularly notable inaccurate belief that drives behavior is pluralistic ignorance. Pluralistic ignorance arises when a majority of individuals privately hold a particular belief, but those individuals believe that most others do not hold that same belief (Miller & McFarland, 1991)⁹. That is, people mistakenly believe that their beliefs are shared by only a minority of others when their beliefs are actually shared by a majority of others. Correcting these mistaken beliefs by informing people that most others share their beliefs can change behaviors. This behavior change can arise by leveraging people's automatic tendency to conform to what they perceive to be the behavior of others (Cialdini, 1993), and also by harnessing social approval or approbation for the behavior. Correcting pluralistic ignorance has been shown to change beliefs and behaviors in a large number of domains, including racial segregation (O'Gorman, 1975), water conservation (Monin & Norton, 2003), gender stereotypes by school children (Prentice & Miller, 1996), alcohol consumption on college campuses (Schroeder & Prentice, 1998), and intergroup contact (Shelton & Richeson, 2005). For example, Schroeder and Prentice (1998) found that many college students (mistakenly) thought that their peers' attitudes toward excessive drinking were more favorable than their own attitudes; reducing this pluralistic ignorance resulted in these students consuming less alcohol than students in a control group four to six months after the intervention had been discontinued. This research illustrates how enduringly changing people's beliefs (in this case, their beliefs about others' attitudes towards excessive drinking) can produce persistent changes in behavior. As another example, Jensen (2010) documented that boys in the Dominican Republic significantly underestimated the returns to education. When these boys were explicitly told the actual returns to education, they completed 0.2 to 0.35 more years of schooling than their peers who did not receive the corrective information.

Interpretation: Interventions that modify how individuals interpret ambiguous information can change the way they respond to future events. This is effectively what cognitive behavioral therapy (CBT) does: it changes how individuals interpret their emotions and thoughts, thereby persistently changing how they respond to their emotions and thoughts (Butler et al., 2006; Beck, 1970). As another example, social adversity (the feeling that one does not belong) can undermine academic motivation and performance (Walton & Cohen, 2007). Walton and Cohen (2011) conducted a study targeting African American students because members of that minority group report feeling socially isolated on many college campuses. When African American students experience feelings of not belonging on campus, many interpret these feelings as unusual and unique to them, exacerbating feelings of being alone and isolated. The intervention guided students to interpret their social adversity as a common and temporary experience that all students—minority and non-minority— go through. Three years after the intervention African

⁹ This arises because others' views are ascertained by public behavior; if others do not publicly espouse the belief that an individual holds, that individual will assume that others do not share his/her belief and will believe that he/she is in the minority (Prentice & Miller, 1993)

American students showed persistent improvement in academic performance and other well-being measures. The persistence of these effects may have been due to the change in interpretation that the intervention prompted. Consistent with this, the persistent treatment effects appeared to be mediated by how students interpreted adversity in their social lives. Other work in education has found that convincing students that their intelligence is based on effort and therefore malleable (as opposed to fixed) can result in higher academic performance over time. These “mindset” interventions can change how students construe future negative performance feedback. The intervention, it is argued, can cause such feedback to be interpreted as signaling the need to work harder rather than as confirmation of students’ beliefs about their own low intelligence (see Dweck, 2007).

Recursive social process pathway. Persistent behavior change may occur if interventions lead other people to enduringly treat the targets in ways that support the behavior change. This can create a reinforcing social cycle, especially if this social process then leads targets to behave differently towards the others. Unlike habit formation and changing mental contents, this pathway is external-to-self because it is driven, at least in part, by the responses of other people. This pathway is similar to the “positive recursive cycles” that Cohen and Garcia (2008) describe. This pathway might also be conceptualized as interventions pushing individuals into new stable social equilibriums (see also Yeager & Walton, 2011).

Interventions targeting student-teacher interactions may have persistent effects due to recursive social cycles. As Gehlbach, Young, and Roan (2012) describe, teachers exposed to interventions targeting student-teacher relationships may treat their students in a more positive manner. This may in turn cause their students to treat their teachers with more respect in future interactions, causing the teacher to respond to the student more positively, and on and on. In this case, the intervention may push the teacher and student into a recursive social process that persists even after the intervention has ended.

Changing future costs pathway. Treatments may cause persistent behavior change if they induce people to perform actions that change the costliness of future behaviors, either financially or in terms of effort. This is another external-to-self pathway, because it induces persistence by leveraging features of the external environment. Sometimes these treatments can make future behaviors less costly (i.e., programming a thermostat today reduces the effort “costs” of future energy conservation) or more costly (i.e., enrolling today in a Gambling Self-Exclusion program—a program in which enrollees are arrested if they enter a casino—increases the future cost of gambling). This pathway may also operate by reducing the *effort* needed to perform a target behavior in the future. Interventions that change future costs or future effort may “lock” people into a trajectory in which either the costs/effort of performing the target behavior are lower (Rogers & Bazerian, 2008), or the costs/effort of failing to perform the target behavior are higher (Bryan et al, 2011). Treatments that lower the future cost of performing the target behavior may reduce the amount of self-control, memory, or resources that are needed to perform a target behavior.

Defaults are one type of intervention that may work through the future costs pathway. As described above, defaulting new employees into 401(k) plans involves a one-time intervention, but it causes a portion of every paycheck employees receive to be automatically deposited into their savings accounts indefinitely into the future (Madrian & Shea, 2001). This savings occurs persistently until the employees proactively (and quite uncommonly) exert effort to change the arrangement. If they do nothing, money will continue to be redirected to their 401(k) accounts. Thus the effort and attention needed to continue saving in the future is lower, suggesting that the persistence of interventions involving defaults may

derive from the way those interventions chance the future costliness of behaviors. This pathway applies to more structural changes that individuals make as well. For example, when people decide in a single moment to invest in energy efficient appliances they gain energy savings every time the appliances are used. Similarly, consider when people program their thermostats, or retrofit their homes: these one-time actions affect future energy usage, without any further effort.

“Rip current” pathway. Another external-to-self pathway through which interventions can generate persistence is through what we term “rip currents.” Rip currents are channels of water in the ocean that run perpendicular to the beach and carry anything that enters them very far into the ocean. People who are just a foot from these channels of water are unaffected by the rip currents; however if individuals move into the channels they could be carried miles out into the ocean by rip currents. By definition, rip currents operate entirely independently of the interventions that are administered and can, when engaged, cause dramatic behavior changes. Interventions can push targets into the on-going current of external processes that reinforce and amplify the interventions’ impacts. These outside processes would not have been engaged if the intervention had not induced the targets to perform initial behaviors. While the processes underlying the rip current pathway may not be strictly psychological, psychologists have discussed them for decades. For example, Kurt Lewin’s notion that people exist in a multi-dimensional tension system in which there are channels that restrain and facilitate each behavior closely relates to the rip current pathway. Lewin’s framework suggests that seemingly minor “channel factors” in situations (social, physical, and cognitive) may open or block channels for dramatic and persistent behavior changes (1946). In other words, external forces can be triggered to compel target behaviors in the future. The social recursive pathway involves targets continuing to engage in and reinforce the social processes that cause persistence (i.e., I smile at you, you smile at me, I smile back, etc.). The rip current pathway, on the other hand, involves people being pushed into streams of on-going influence that are external to the intervention and do not require target individuals to experience continued social interactions.

In get-out-the-vote research, a common finding is that inducing people to vote in one election leads to greater turnout in subsequent elections many years later (Gerber, Green, and Shachar, 2003; Davenport et al., 2010). One factor that may contribute to this is that once people have voted in one election (and the publicly available administrative records report that they have), future campaigns target them differently and more intensively. An intervention that induces people to vote in a given election might push them into the “current” of future campaign activity. In this scenario, in future elections the individuals may become targets of increased voter mobilization activity (phone calls, mailings, etc.). This would mean that the original intervention pushed targets into the current of subsequent “interventions” that have no connection to the original intervention.

Conclusion

In this manuscript we explore the behavioral science of how interventions work over time. We discussed the features that influence whether interventions can bridge the gap between when they are administered and when the target behaviors are to be performed, and the features that affect whether these interventions generate marginal benefits to repeated treatment. We also discussed the pathways through which interventions can generate persistent treatment effects after the interventions have been discontinued. The range of disciplines from which we have cited—health, energy conservation, education, marketing, politics, and consumer behavior— illustrates the importance of this topic across the social sciences.

Our aim was both to develop frameworks that have utility to researchers and practitioners, and to highlight areas where more research is needed. For example, among the handful of studies that have documented long-term treatment effects, many (e.g. Rodin & Langer, 1977; Fryer et al., 2012) involve research designs that conflate marginal benefits to continued treatment and persistence. As we have discussed here, these are different phenomena and future research should treat them as such. Future research could be structured so as to disentangle these concepts by varying the number of times a behavioral intervention is repeated, in addition to measuring how long the effects can be detected after the intervention has been discontinued.

One open question of particular importance is what moderates the persistence of treatment effects after treatments have been discontinued? This question is highlighted by recent work by Allcott and Rogers (in press). They report three field experiments in which households received for several years, on a monthly or quarterly basis, mailings that compared targets' energy usage to that of their neighbors. This treatment is administered by the energy efficiency company OPOWER. There were three conditions. Households in the control condition did not receive any mailers; households in the continued condition received mailers continually over the multi-year study period; households in the discontinued condition received the mailers for only the first half of the study period. The researchers report three findings. First, those in both the continued and discontinued conditions used less energy than those in the control condition. Given that energy usage decisions (turning off the lights, programming the thermostat, buying CFL lightbulbs, etc.) likely occur temporally delayed from when people read their mail, this means that the intervention successfully generated an impact by bridging time. Second, after the mailers were no longer sent to those in the discontinued treatment condition those in the continued condition used less energy than those in the discontinued treatment condition. This is evidence of the marginal benefit to continued treatment. Third, several years after the mailings were discontinued those in the discontinued condition used less energy than those in the control condition. This is evidence of persistence. Tables 1-3 summarize the three sections of this manuscript and illustrate each feature and pathway with speculation about how these processes might affect the experiments reported by Allcott and Rogers.

An intriguing finding across these experiments is that the degree of persistence varied meaningfully across contexts. This means that the degree of persistence (and the related rate of decay once the treatment is discontinued) of the same treatment varied across sites. Why? At this point we do not know, but we anticipate future research will explore this and related questions.

Behavioral science is increasingly being used to develop interventions to influence important behaviors throughout society. From firms to governments, from community organizers to teachers, behavioral science offers insight and tools for changing behavior. With this manuscript we sought to show that there are knowable aspects of behavioral interventions that affect whether, how, and for how long they change behavior. Our hope is that scholars find these frameworks productive for advancing and organizing future research, and that they help those who develop behavioral interventions to make them more effective.

REFERENCES

- Achtziger, A., Gollwitzer, P.M., & Sheeran, P. (2008). Implementation Intentions and Shielding Goal Striving From Unwanted Thoughts and Feelings. *Personality and Social Psychology Bulletin*, *34*, 381-392.
- Acland, D., & Levy, M. (2010). Habit formation and Naivete in gym attendance: evidence from a field experiment. *Unpublished paper*.
- Allcott, H., R. & Rogers, T. (in press). The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. *American Economic Review*
- Amir, O., Ariely, D., Cooke, A., Dunning, D., Epley, N., Gneezy, U., Koszegi, B., Lichtenstein, D., Mazar, N., Mullainathan, S., Prelec, D., Shafir, E., & Silva, J. (2005). Psychology, Behavioral Economics, and Public Policy. *Marketing Letters*, *16*(3), 443-454.
- Austin, J., Sigurdsson, S.O., & Rubin, Y.S. (2006). An Examination of the Effects of Delayed Versus Immediate Prompts on Safety Belt Use. *Environment and Behavior*, *38*(1), 140-149.
- Beck, A.T. (1970). Cognitive Therapy: Nature and Relation to Behavior Therapy. *Behavior Therapy*, *1*(2), 184-200.
- Bryan, C. J., Walton, G. M., Rogers, T., & Dweck, C. S. (2011). Motivating voter turnout by invoking the self. *Proceedings of the National Academy of Sciences*, *108*, 12653-12656.
- Bryan, G., Karlan, D., & Nelson, S. (2010). Commitment Devices. *Annual Review of Economics*, *2*, 671-698.
- Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clinical psychology review*, *26*(1), 17-31.
- Cahill, K., & Perera, R. (2008). Competitions and incentives for smoking cessation. *Cochrane Database of Systematic Reviews*, *3*.
- Camerer, C., Loewenstein, G., & Prelec, D. (2003). Neuroeconomics: How Neuroscience Can Inform Economics. Caltech Working Paper.
- Charness, G., & Gneezy, U. (2009). Incentives to exercise. *Econometrica*, *77*(3), 909-931.
- Chattopadhyay, R., & Duflo, E. (2004). Women as Policy Makers: Evidence from a Randomized Policy Experiment in India. *Econometrica*, *72*(5), 1409-1443.
- Chugh, D. & Bazerman, M.H. (2007). Bounded awareness: what you fail to see can hurt you. *Mind & Society*, *6*, 1-18.
- Cialdini, R. B., Borden, R. J., Thorne, A., Walker, M. R., Freeman, S., & Sloan, L. R. (1976). Basking in reflected glory: Three (football) field studies. *Journal of Personality and Social Psychology*, *34*(3), 366.
- Cialdini, R. B. (1993). Influence: The psychology of persuasion.
- Cohen, G.L., & Garcia, J. (2008). Identity, belonging, and achievement: A model, interventions, implications. *Current Directions in Psychological Science*, *17*(6), 365-369.

- Davenport, T.C., Gerber, A.S., Green, D.P., Larimer, C.W., Mann, C.B., & Panagopoulos, C. (2010). The Enduring Effects of Social Pressure: Tracking Campaign Experiments Over a Series of Elections. *Political Behavior*, 32, 423-430.
- Della Vigna, S. & Malmendier, U. (2006). Paying not to go to the gym. *American Economic Review*, 96(3), 694-719.
- Detweiler, J.B., Bedell, B.T., Salovey, P., Pronin, E., & Rothman, A.J. (1999). Message Framing and Sunscreen Use: Gain-Framed Messages Motivate Beach-Goers. *Health Psychology*, 18(2), 189-196.
- Dweck, C. (2007). *Mindset: The new psychology of success*. Random House Publishing, New York.
- Ellis, N. C. (1996). Analyzing language sequence in the sequence of language acquisition: Some comments on Major and Ioup. *Studies in Second Language Acquisition*, 18(03), 361-368.
- Ferraro, P.J., Miranda, J.J., & Price, M.K. (2011). The Persistence of Treatment Effects with Norm-Based Policy Instruments: Evidence from a Randomized Environmental Policy Experiment. *American Economic Review: Papers and Proceedings*, 101(3), 318-322.
- Freedman, J.L., & Fraser, S.C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology*, 4(2), 195-202.
- Friestad, M., & Wright, P. (1994). The Persuasion Knowledge Model: How People Cope with Persuasion Attempts. *Journal of Consumer Research*, 21, 1-31.
- Fryer, R.G., Levitt, S.D., List, J., & Sadoff, S. (2012). Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment. NBER Working Paper No. 18237. The National Bureau of Economic Research.
- Ganzach, Y., & Karsahi, N. (1995). Message framing and buying behavior: a field experiment. *Journal of Business Research*, 32(1), 11-17.
- Gehlbach, H., Young, L.V., & Roan, L.K. (2012). Teaching social perspective taking: how educators might learn from the Army. *Educational Psychology*, 1-15.
- Gerber, A.S., Green, D.P., & Larimer, C.W. (2008). Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment. *American Political Science Review*, 102(1), 33-48.
- Gerber, A., Green, D., & Shachar, R. (2003). Voting may be habit-forming: Evidence from a randomized field experiment. *American Journal of Political Science*, 47(3), 540-550.
- Gerber, A.S., Huber, G.A., Doherty, D., Dowling, C.M., & Hill, S.J. (2011). Do Perceptions of Ballot Secrecy Influence Turnout? Results from a Field Experiment. NBER Working Paper No. 17673.
- Gilovich, T., Griffin, D., & Kahneman, D., eds. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press, Cambridge, U.K.
- Gollwitzer, P.M. (1999). Implementation Intentions: Strong Effects of Simple Plans. *American Psychologist*, 54(7), 493-503.
- Gollwitzer, P.M., & Sheeran, P. (2006). Implementation Intentions and Goal Achievement: A Meta-analysis of Effects and Processes. *Advances in Experimental Social Psychology*, 38, 69-119.

- Hallam, J. S., & Petosa, R. (2004). The long-term impact of a four-session work-site intervention on selected social cognitive theory variables linked to adult exercise adherence. *Health Education & Behavior, 31*(1), 88-100.
- Harris, J.L., Bargh, J.A., & Brownell, K.D. (2009). Priming Effects of Television Food Advertising on Eating Behavior. *Health Psychology, 28*(4), 404-413.
- Jensen, M. C. (2010). Value maximization, stakeholder theory, and the corporate objective function. *Journal of applied corporate finance, 22*(1), 32-42.
- John, L., Loewenstein, G., Troxel, A., Norton, L., Fassbender, J., & Volpp, K. (2011). Financial incentives for extended weight loss: A randomized, controlled trial. *Journal of General Internal Medicine, 26*(6), 621-626.
- Johnson, E.J., & Goldstein, D. (2003). Do Defaults Save Lives? *Science, 302*, 1338-1339.
- Johnson, E. J., Häubl, G., & Keinan, A. (2007). Aspects of endowment: a query theory of value construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 461.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263-291.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological bulletin, 108*(3), 480.
- Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics, 112*(2), 443-478.
- LeBoeuf, R.A., Shafir, E., & Bayuk, J.B. (2010). The conflicting choices of alternating selves. *Organizational Behavior and Human Decision Processes, 111*(1), 48-61.
- Lewin, K. (1946). Behavior and development as a function of the total situation. L. Carmichael, ed. *Manual of Child Psychology*, John Wiley & Sons Inc., Hoboken, NJ, 791-844.
- List, J.A. (2011). Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off. *Journal of Economic Perspectives, 25*(3), 3-16.
- Lord, Ross, and Lepper, 1979
- Madrian, B., Shea, D. (2001). The power of suggestion: Inertia in 401(k) participation and savings behavior. *Quarterly Journal of Economics, 116*(4), 1149-1187.
- Mataconis, D. 2012. Is It Anyone's Business If You Don't Vote? Outside the Beltway. 2 June 2012. Available online at < <http://www.outsidethebeltway.com/is-it-anyones-business-if-you-dont-vote/>>
- McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied cognitive psychology, 14*(7), S127-S144.
- McGuire, W.J. (1964). Inducing resistance to persuasion: Some contemporary approaches. *Advances in Experimental Social Psychology, 1*, 191-229.
- McSweeney, F.K. (2004). Dynamic Changes in Reinforcer Effectiveness: Satiation and Habituation Have Different Implications for Theory and Practice. *Behavior Analyst, 27*(2), 171-188.

- Milkman, K.L., Beshears, J., Choi, J.J., Laibson, D., & Madrian, B.C. (2011). Using implementation intentions prompts to enhance influenza vaccination rates. *PNAS*, *108*(26), 10415-10420.
- Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2012). Following through on good intentions: The power of planning prompts. NBER Working Paper No. 17995. Cambridge, MA: National Bureau of Economic Research.
- Milkman, K., Minson, J., & Volpp, K. (2012). Holding the Hunger Games Hostage at the Gym: An Evaluation of Temptation Bundling. *The Wharton School Research Paper*, 45.
- Milkman, K. L., Rogers, T., & Bazerman, M. H. (2008). *Highbrow films gather dust: A study of dynamic inconsistency and online DVD rentals* (No. 07-099). Harvard Business School.
- Miller, D.T., & McFarland, C. (1991). When social comparison goes awry: The case of pluralistic ignorance. J. Suis, T. Ashby, ed. *Social Comparison: Contemporary Theory and Research*, Lawrence Erlbaum Associates, Hillsdale, NJ, 287-313.
- Milne, S., Orbell, S., & Sheeran, P. (2002). Combining motivational and volitional interventions to promote exercise participation: Protection motivation theory and implementation intentions. *British Journal of Health Psychology*, *7*(2), 163-184.
- Monin, B., & Norton, M.I. (2003). Perceptions of a fluid consensus: Uniqueness bias, false consensus, false polarization, and pluralistic ignorance in a water conservation crisis. *Personality and Social Psychology Bulletin*, *29*(5), 559-567.
- Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as a limited resource: Regulatory depletion patterns. *Journal of personality and social psychology*, *74*(3), 774.
- Nickerson, D.W. (2007). Quality is Job One: Professional and Volunteer Voter Mobilization Calls. *American Journal of Political Science*, *51*(2), 269-282.
- Nickerson, D., & Rogers, T. (2010). Do you have a voting plan? Implementation intentions, voter turnout, and organic plan making. *Psychological Science*, *21*(2), 194-199.
- O’Gorman, H.J. (1975). Pluralistic ignorance and White estimates of White support for racial segregation. *Public Opinion Quarterly*, *39*(3), 313-330.
- Ouellette, J., & Wood, W. (1998). Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychological Bulletin*, *124*(1), 54-74.
- Petty, R. E. (2013). Two routes to persuasion: State of the art. *International perspectives on psychological science*, *2*, 229-247.
- Petty, R.E., & Cacioppo, J.T. (1986). *Communications and persuasion: Central and peripheral routes to attitude change*. New York: Springer-Verlag.
- Petty, R. E., & Cacioppo, J. T. (1977). Forewarning, cognitive responding, and resistance to persuasion. *Journal of Personality and Social Psychology*, *35*(9), 645.
- Prentice, D.A. & Miller, D.T. (1993). Pluralistic Ignorance and Alcohol Use on Campus: Some Consequences of Misperceiving the Social Norm. *Journal of Personality and Social Psychology*, *64*(2), 243-256.

- Prentice, D.A. & Miller, D.T. (1996). Pluralistic Ignorance and the Perpetuation of Social Norms by Unwitting Actors. *Advances in Experimental Social Psychology*, 28, 161-209.
- Rankin, C.H., Abrams, T., Barry, R.J., Bhatnagar, S., Clayton, D., Colombo, J., Coppola, G., Geyer, M.A., Glanzman, D.L., Marsland, S., McSweeney, F., Wilson, D.A., Wu, C.F., & Thompson, R.F. (2009). Habituation Revisited: An Updated and Revised Description of the Behavioral Characteristics of Habituation. *Neurobiology of Learning and Memory*, 92(2), 135-138.
- Rodin, J., & Langer, E.J. (1977). Long-Term Effects of a Control-Relevant Intervention With the Institutionalized Aged. *Journal of Personality and Social Psychology*, 35(12), 897-902.
- Rogers, T., & Bazerman, M. H. (2008). Future lock-in: Future implementation increases selection of 'should' choices. *Organizational Behavior and Human Decision Processes*, 106(1), 1-20.
- Rogers and Ternovski (2014). "Look out for our call after the election": Who responds to the threat of reporting one's own behavior? *Working paper*.
- Ross, L., & Nisbett, R. (1991). *The Person and the Situation: Perspectives on social psychology*. McGraw Hill, New York.
- Rothman, A. J., Sheeran, P., & Wood, W. (2009). Reflective and automatic processes in the initiation and maintenance of dietary change. *Annals of Behavioral Medicine*, 38(1), 4-17.
- Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American psychologist*, 54(3), 182.
- Schank, R. C., Abelson, R. P., & Scripts, P. (1977). Goals and understanding. *Erlbaum: Eksevier Science*.
- Schroeder, C. M., & Prentice, D. A. (1998). Exposing Pluralistic Ignorance to Reduce Alcohol Use Among College Students. *Journal of Applied Social Psychology*, 28(23), 2150-2180.
- Schultz, P., Nolan, J., Cialdini, R., Goldstein, N., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18(5), 429-434.
- Schwartz, J., Mochon, D., Wyper, L., Maroba, J., Patel, D., & Ariely, D. (2014). Healthier by Precommitment. *Psychological Science*, 538-546.
- Shah, A. K., Mullainathan, S., & Shafir, E. (2012). Some consequences of having too little. *Science*, 338(6107), 682-685.
- Shelton, J.N., & Richeson, J.A. (2005). Intergroup Contact and Pluralistic Ignorance. *Journal of Personality and Social Psychology*, 88(1), 91-107.
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception-London*, 28(9), 1059-1074.
- Sunstein, C. R. (2013). Office of Information and Regulatory Affairs: Myths and Realities, The.
- Sweller, J., Van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, 10(3), 251-296.
- Thaler, R., & Sunstein, C. (2003). Libertarian Paternalism. *American Economic Review*, 92(2), 175-179.

- Thaler, R., & Sunstein, C. (2009). *Nudge: Improving decisions about health, wealth and happiness*. Penguin, New York.
- Thompson, R.F., & Spencer, W.A. (1966). Habituation: A Model Phenomenon for the Study of Neuronal Substrates of Behavior. *Psychological Review*, 73, 16-43.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453-458.
- Volpp, K., Troxel, A., Pauly, M., Glick, H., Puig, A., Asch, D., Galvin, R., Zhu, J., Wan, F., deGuzman, J., Corbett, E., Weiner, J., & Audrain-McGovern, J. (2009). A randomized, controlled trial of financial incentives for smoking cessation. *New England Journal of Medicine*, 360(February 12), 699-709.
- Walton, J., & Cohen, G. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology*, 92(1), 82-96.
- Walton, J., & Cohen, G. (2011). A brief social-belonging intervention improves academic health outcomes of minority students. *Science*, 331(6023), 1447-1451.
- Wood, W., Tam, L., & Witt, M. G. (2005). Changing circumstances, disrupting habits. *Journal of Personality and Social Psychology*, 88(6), 918.
- Yeager, D., & Walton, G. (2011). Social-psychological interventions in education. *Rev. Educational Research*, 81(2), 267-301.

TABLE 1

Features Likely to Bridge Time	
Feature	Speculation about How Feature Might Affect OPOWER Energy Efficiency Treatment Effect
Interventions that Chronically Direct Bounded Attention	People feel that they are being held socially accountable for their energy usage so it becomes particularly cognitively accessible after receiving the reports
Interventions that Link a Performance Environment with an Intention	People may make a plan for how they will reduce their electricity use; for example, they make a plan to turn off the air conditioning as they leave for work in the morning. This results in people remembering as they walk out the door in the morning that they need to turn off their air conditioners
Interventions that Allow Individuals to Pre-commit to a Future Behavior	People may pre-commit that if they use more than a certain amount of electricity that they will donate a given amount of money to an organization they oppose
Interventions that Enduringly Change Consequential Thoughts	People become aware that other people are more energy efficient than they are, whereas they had not known that to be the case. Enduringly changing their perception of the descriptive norm prompts them to change their energy usage behavior

TABLE 2

Features Likely to Produce Marginal Benefits to Continued Treatment	
Feature	Speculation about How Feature Might Affect OPOWER Energy Efficiency Treatment Effect
Interventions that Resist Attention Habituation	People receive monthly reports that have dynamic content each time. In this way, the reports are unique each time and prevent habituation.
Interventions that Produce Incomplete or Temporary Change in Behavior	People replace their appliances, light bulbs, windows, and other energy-intensive possessions gradually over time. Repeated treatment can affect whatever choices are imminent.
Interventions that Induce Minimal Resistance or Distrust	People receive the reports and understand that the reports aim to influence recipients' energy choices. The directness and credibility of the messenger (e.g., the utility) decrease likelihood of distrust.

TABLE 3

Pathways to Persistence		
Pathway	Description	Speculation about How Feature Might Affect OPOWER Energy Efficiency Treatment Effect
Habit Pathway	Treatment produces an automatic tendency to repeat a particular behavioral response, triggered by a stable context in which the behavior is performed	People begin to consciously turn off the lights every time they leave a room; eventually the contextual cue (exiting the room) automatically triggers the behavior (turning off the lights)
Changes to Mental Contents Pathway	Treatment permanently changes beliefs or interpretations that are causally consequential to target behaviors. This may arise by replacing existing beliefs with different beliefs, by creating beliefs where they did not previously exist, or by changing the way people interpret ambiguous stimuli	<p>People open their windows instead of using the air conditioners, and they learn that they enjoy fresh air in their homes even when it is warmer than they are used to (results in a change in preferences)</p> <p>After taking energy-saving steps, people develop a self-perception that they are the kinds of people who save energy (“energy savers”), which makes them behave in self-perception consistent ways in the future (i.e., turning the lights off when they leave rooms and purchasing energy efficient products)</p>
Recursive Social Processes Pathway	Treatment leads other people to enduringly treat targets in ways that support the behavior change	People purchase energy saving products like solar panels and begin showing their friends, who then regularly ask the people about their energy use, effectively putting continued social pressure on the people to continue saving energy
Future Costs Pathway	Treatment induces people to perform behaviors that change the costliness of future behaviors; the treatment may decrease cost of performing target behaviors, or increase cost of failing to perform target behaviors	<p>People make a one-time decision to program their “smart” thermostats, which reduces their electricity usage in the future</p> <p>People retrofit their homes so their homes require less energy</p> <p>People purchase energy efficient appliances</p>
Rip Currents Pathway	Treatment induces people to perform a behavior that then exposes them to on-going external processes that they would not have been exposed to otherwise; these external processes cause the changed behavior to persist	<p>People buy an energy-efficient appliance and are added to marketing lists for other energy efficiency products, which they also buy and which remind the people that they need to save more energy</p> <p>Organizations like the Sierra Club target people based on these marketing lists, which engages them in ways that further reinforce energy saving behaviors</p>

