Journal of Applied Psychology

A Field Experiment on Subgoal Framing to Boost Volunteering: The Trade-Off Between Goal Granularity and Flexibility

Aneesh Rai, Marissa A. Sharif, Edward H. Chang, Katherine L. Milkman, and Angela L. Duckworth

Online First Publication, September 15, 2022. http://dx.doi.org/10.1037/apl0001040

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Applied Psychology

ISSN: 0021-9010

https://doi.org/10.1037/apl0001040

A Field Experiment on Subgoal Framing to Boost Volunteering: The Trade-Off Between Granularity and Flexibility

Aneesh Rai[1], Marissa A. Sharif[2], Edward H. Chang[3], Katherine L. Milkman[1], and Angela L. Duckworth[1]

[1] Department of Operations, Information and Decisions, The Wharton School, University of Pennsylvania

[2] Department of Marketing, The Wharton School, University of Pennsylvania

[3] Department of Negotiation, Organizations and Markets, Harvard Business School, Harvard University

Research suggests that breaking overarching goals into more granular subgoals is beneficial for goal progress. However, making goals more granular often involves reducing the flexibility provided to complete them, and recent work shows that flexibility can also be beneficial for goal pursuit. We examine this trade-off between granularity and flexibility in subgoals in a preregistered, large-scale field experiment ($N = 9{,}108$) conducted over several months with volunteers at a national crisis counseling organization. A preregistered vignette pilot study ($N = 900$) suggests that the subgoal framing tested in the field could benefit goal seekers by bolstering their self-efficacy and goal commitment, and by discouraging procrastination. Our field experiment finds that reframing an overarching goal of 200 hr of volunteering into more granular subgoals (either 4 hr of volunteering every week or 8 hr every 2 weeks) increased hours volunteered by 8% over a 12-week period. Further, increasing subgoal flexibility by breaking an annual 200-hr volunteering goal into a subgoal of volunteering 8 hr every 2 weeks, rather than 4 hr every week, led to more durable benefits.

Keywords: goals, subgoals, field experiment, flexibility

Goals vary on many dimensions, including their granularity. For example, one could commit to a volunteering goal of 200 hr in a year. However, that same overall goal could be broken down into more granular subgoals, such as committing to volunteering 8 hr every 2 weeks or 4 hr every week for a year. Increasingly granular subgoals have the benefit of breaking large targets down into more manageable parts, but they inherently have less flexibility (i.e., they allow for fewer possible ways of achieving a goal). In this article, we explore the trade-off between increasing subgoal flexibility and subgoal granularity in a preregistered, longitudinal field experiment with thousands of volunteer crisis counselors.

Dividing goals into more granular subgoals may be beneficial for several reasons. First, past research has found that it can increase self-efficacy to achieve subgoals, in turn making overarching goals seem more attainable (Bandura & Schunk, 1981; Latham & Seijts, 1999). In addition, dividing goals into more granular subgoals may reduce procrastination by creating more frequent and imminent deadlines (Ariely & Wertenbroch, 2002; Janakiraman & Ordóñez, 2012; Lieberman et al., 2021). Finally, it can increase commitment to overarching goals by helping people focus on making small, near-term sacrifices of time, which are less daunting than large, distant sacrifices (Gourville, 1998; Hershfield et al., 2020). As subgoals become more granular, each of these benefits should be magnified.

However, breaking large goals down into specific subgoals also means reducing the amount of flexibility available to a goal seeker in terms of how the goal is achieved. Reduced flexibility means greater chances for goal failure or other setbacks, which have been shown to increase the risk of goal abandonment (Cochran & Tesser, 1996; Soman & Cheema, 2004; cf. Bandura & Locke, 2003). In addition, allowing goal seekers to pursue their goals more flexibly has been shown to have various benefits, such as increasing control over scheduling—which can boost well-being—as well as improving performance and elevating persistence in the face of failure to achieve a goal (Beshears et al., 2021; Moen et al., 2016; Sharif & Shu, 2017, 2021).

We examine the trade-off between the granularity and flexibility of subgoals in a large, preregistered field experiment in which we assess the effects of framing the same goal differently on objective hours of volunteering on a crisis counseling platform over a 3-month

Aneesh Rai https://orcid.org/0000-0001-9633-2568

The authors would like to thank Freddie Bologno, Bob Filbin, Shannon Green, Elise Segars, Tessa Shapiro, Elizabeth Sweezey, and Jaclyn Weiser, some of our many contacts at Crisis Text Line who supported this work. The authors also thank Christophe Van den Bulte and Lorin Hitt for their early feedback on this work. The authors also thank Katherine Chan, Meghan Chung, and Timothy Lee for their research assistance; Kasandra Brabaw and Michelle Shih for editorial input; and Amaris Kobolak for graphic design. Earlier versions of this work were presented at the 2020 Center for Health Incentives & Behavioral Economics-Palliative and Advanced Illness Center (CHIBE-PAIR) Roybal Mini-Symposium, the 2021 Society for Personality and Social Psychology Judgment and Decision Making Pre-Conference, the 2021 East Coast Doctoral Conference, the 2021 Association for Consumer Research Conference, the 2021 Society for Judgment and Decision Making Conference, the 2021 Massachusetts Institute of Technology Conference on Digital Experimentation, and the 2022 Behavioral Science and Policy Association Annual Conference.

Correspondence concerning this article should be addressed to Aneesh Rai, Department of Operations, Information and Decisions, The Wharton School, University of Pennsylvania, 532.1 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104, United States. Email: aneeshr@wharton.upenn.edu

period. Specifically, we vary whether a volunteering goal is framed as "200 hr in a year" (least granular and most flexible) or whether that goal is broken down into increasingly granular but less flexible subgoals—either volunteering "8 hr every 2 weeks" (more granular and less flexible) or "4 hr every week" (even more granular and even less flexible). We also present results from a preregistered, online pilot study mirroring our field experiment and, as theorized, we find that subgoal framing influences forecasted self-efficacy, procrastination, and goal commitment.

Our article makes several key contributions to the goal-setting literature. First, previous work examining the benefits of subgoals has largely examined one-time decisions studied in the laboratory (Amir & Ariely, 2008; Fishbach et al., 2006; Latham & Seijts, 1999; Seijts & Latham, 2001; Stock & Cervone, 1990), and the few existing field studies have included 70 or fewer participants per condition, raising concerns about statistical power (Bandura & Schunk, 1981; Bandura & Simon, 1977; Huang et al., 2017; Latham & Brown, 2006). Here, we present a well-powered (over 3,000 participants per condition) and ecologically valid examination of the value of subgoals in a field context where we measure objective levels of goal progress over time. Further, we explore the trade-off between more granular versus more flexibly framed subgoals. We also expand on previous theory by highlighting several possible benefits of subgoals that have previously been overlooked—namely, that they may reduce procrastination by creating more imminent deadlines, and that they may increase goal commitment by requiring smaller, near-term sacrifices of time.

Theoretical Foundations and Hypotheses

The Benefits of Granular Subgoals

Previous laboratory research suggests that breaking a large goal down into more granular subgoals can be beneficial for goal pursuit (Amir & Ariely, 2008; Latham & Seijts, 1999; Stock & Cervone, 1990). For example, in one study, participants took part in a complex, multiround business simulation where goal-relevant information changed across rounds (Latham & Seijts, 1999). Those who were encouraged to try to earn a certain amount of money in each round (i.e., those assigned granular subgoals) as well as to pursue an overall earnings goal (i.e., those assigned a less granular overarching goal) outperformed their peers who were only provided with an overarching goal. These findings have been extended to small field studies examining outcomes such as weight loss, academic performance among MBA students, arithmetic performance among elementary school students, and the number of photos taken for a market intelligence task (Bandura & Schunk, 1981; Bandura & Simon, 1977; Huang et al., 2017; Latham & Brown, 2006).

Why are subgoals beneficial? Past research has posited that they increase self-efficacy by providing early markers of accomplishment and making distal goals seem more attainable (Bandura & Schunk, 1981; Latham & Seijts, 1999). Self-efficacy is generally theorized to be beneficial for goal pursuit (Bandura & Schunk, 1981; Locke & Latham, 2019; cf. Vancouver et al., 2001). For example, Latham and Seijts (1999) found that assigning people granular subgoals (but not less granular overarching goals) in a laboratory task led to an increase in participants' self-efficacy, which in turn was correlated with better performance.[1]

In addition, we propose that more granular subgoals may discourage procrastination. The rewards for achieving overarching goals (i.e., goal completion) are in the distant future, and people have a well-established tendency to impatiently choose smaller, short-term rewards over larger, long-term rewards (Ainslie, 1975; Schouwenburg & Groenewoud, 2001; Steel, 2007). This often leads to procrastination, which is a key obstacle to goal initiation and completion (Krause & Freund, 2014). The risk of procrastination may be reduced by more granular subgoals because subgoals yield more immediate consequences. Specifically, breaking an overarching goal into a series of more granular subgoals produces more frequent and immediate deadlines, and more frequent and immediate deadlines help combat procrastination (Ariely & Wertenbroch, 2002; Janakiraman & Ordóñez, 2012; Lieberman et al., 2021; Zhu et al., 2019). For example, Ariely and Wertenbroch (2002) found that participants who were assigned subgoals in the form of three intermediate deadlines for different proofreading tasks were more proficient at their work than participants who were simply assigned an overarching deadline for all assignments.

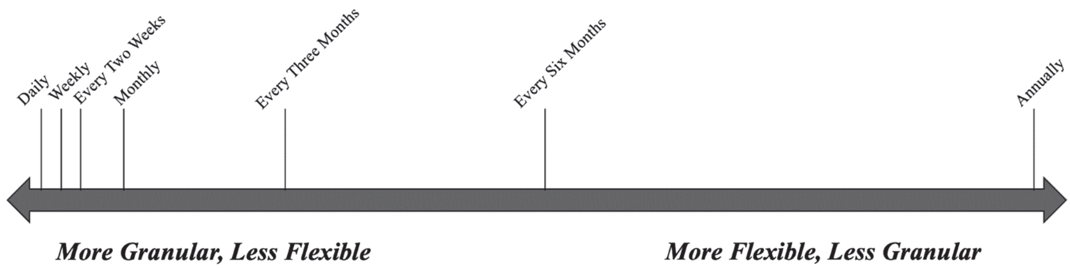
Another way we theorize that more granular subgoals may improve goal progress is by making it easier for people to commit to goals because, by design, more granular subgoals require smaller commitments of time than overarching goals. Recent research has shown the benefits of reframing large goals by describing them in terms of subgoals that are perceived as smaller commitments (Gourville, 1998; Hershfield et al., 2020). In one study, Hershfield et al. (2020) found that consumers were 45% more likely to sign up for a savings program framed as deducting \$35 from their bank account weekly than an identical program framed in a less granular way as deducting \$150 from their bank account monthly. However, framing the savings program to be even more granular—deducting \$5 a day from their account—led to the most signups, with a fourfold increase relative to the monthly condition and an almost threefold increase relative to the weekly condition. Hershfield et al. theorized that consumers found it less psychologically painful to give up smaller amounts of money at a higher frequency than an equivalent one-time lump sum. While this research examined a one-time, effortless decision, we argue that even in the domain of effortful goal pursuit, people may find it easier to commit to a series of more granular subgoals rather than a single, overarching goal.

The Risks of More Granular Subgoals and the Benefits of Flexibility

While more granular subgoals have a number of theorized benefits, they also come with several risks. First, more granular goals generally provide less flexibility to goal seekers regarding ways to accomplish their goal. This is always the case when goal granularity is achieved by framing overarching goals in terms of a series of smaller, more temporally proximal subgoals, which is the tactic we study (see Figure 1). For example, there are objectively

[1] Past work has also theorized that subgoals encourage people to try new strategies and make errors quickly. This can increase the frequency of feedback and facilitate the discovery of successful goal-pursuit strategies, particularly for complex tasks (Bandura & Simon, 1977; Frese & Zapf, 1994; Latham & Seijts, 1999). However, past theory would not predict that this mechanism would apply to settings like the one we study, where the task is simply logging into an online platform to complete committed hours of work.

Figure 1
Goals Lie on a Continuum From More Granular to Less Granular



Note. When goal granularity is achieved by framing overarching goals in terms of a series of smaller, more temporally proximal subgoals, those subgoals become less flexible.

more possible ways to accomplish an annual goal of visiting the gym 120 times than ways to achieve a year-long goal of making 10 gym visits every month.

A key reason flexibility can be helpful to anyone pursuing an overarching goal is that it reduces the negative consequences of goal violations, which can lead people to completely give up on their overarching goal. Research on the "what-the-hell effect" has demonstrated that goal violations increase the chances of goal abandonment (Cochran & Tesser, 1996). For example, Herman and Mack (1975) found that dieters who went over their daily calorie limit (i.e., a goal violation) often ended up overindulging (i.e., abandoning goal progress entirely). Work by Soman and Cheema (2004) also found that when assigning participants to a proofreading task with either standard goals, subgoals, or no goals, participants who failed to reach their assigned goal performed worse than those who never set goals because goal failures were so demotivating.[2]

Recent work has demonstrated the benefits of flexibility for goal pursuit and buffering against the demotivating effects of goal failures. One study by Sharif and Shu (2021) demonstrated the benefits of incorporating psychological flexibility into goals by giving people a way to avoid perceiving a misstep as a goal failure. Framing goals with "emergency reserves" (e.g., a goal of going to the gym 7 days of the week with 2 "emergency" skip days) improved goal performance by reducing people's sense that their goal progress had been interrupted by a violation. In another study, Beshears et al. (2021) found that encouraging aspiring gym-goers to build an exercise habit on a flexible schedule (making gym visits at variable times) for a month led people to exercise significantly more often in the long run than encouraging people to build a more rigid exercise habit (making most gym visits at the same time of day). In other words, more possible paths to success simply produced better outcomes. Finally, flexibility in goal pursuit increases schedule control (i.e., the ability to decide when to work), and schedule control is associated with various benefits, including greater well-being and work–life balance (Kelly et al., 2011; Moen et al., 2016).

To summarize, more granular subgoals of the type we study may improve goal progress by bolstering self-efficacy and goal commitment and by reducing procrastination. However, as goals are broken down into more and more granular subgoals, they become less flexible, which—in the extreme—may end up hampering success, producing an inverted U-shaped relationship between goal granularity and goal progress such that more granular subgoals are beneficial only up to a point.

There is another risk of breaking large goals down into more granular subgoals: breeding a sense of complacency. Subgoals provide salient reference points as well as early markers of accomplishment (Bandura & Schunk, 1981; Heath et al., 1999). Feelings of accomplishment can boost self-efficacy, which is generally considered beneficial for goal pursuit (Bandura & Schunk, 1981; Locke & Latham, 2019), but high self-efficacy can also undermine goal pursuit by promoting a sense of complacency once a subgoal is achieved (Fishbach et al., 2006; Vancouver et al., 2001, 2002). Laboratory studies corroborate the notion that subgoals can sometimes backfire by promoting a sense of complacency (Amar et al., 2011; Amir & Ariely, 2008; Fishbach et al., 2006). However, Gal and McShane (2012) theorize that while people might focus on subgoals in the short run, their attention should refocus on their overarching goal over time, which should reduce the risks of goal complacency produced by more granular subgoals over long time horizons like in our field experiment.

Hypotheses

In this article, we test the effects of reframing an overarching goal as a series of smaller, more temporally proximal subgoals, and we vary the granularity (and therefore flexibility) of those subgoals. Past theory and research suggest that breaking a large, overarching goal down into a series of more granular subgoals should be generally beneficial, even with the loss of flexibility that accompanies this shift. Therefore, we hypothesize the following:

> *Hypothesis 1:* Reframing an overarching goal as a series of more granular subgoals (e.g., focusing on objectives to accomplish "every 2 weeks" or "every week" instead of "this year") will increase goal progress.

However, goal flexibility is also an asset. We, therefore, propose a curvilinear relationship between goal granularity and goal progress such that when goals become too granular—and therefore, inflexible—the benefits of subgoals diminish and may reverse.

[2] Notably, goal failures do not always harm future performance, and when goal failure is harmful has been theorized to depend on factors such as self-efficacy, self-dissatisfaction with the goal failure, and level of goal failure (Bandura & Cervone, 1986).

Hypothesis 2: There is a limit to the benefits of making subgoals more granular (e.g., focusing on objectives to accomplish "every week," "every day," or "every hour"); excessively granular (and inflexible) subgoals will cease to improve goal progress and will instead harm it.

Finally, the benefits of goal flexibility may be particularly important over the long run because opportunities for goal failure tend to accumulate over time (Norcross & Vangarelli, 1988–1989; Oscarsson et al., 2020). For example, Oscarsson et al. (2020) found that people become increasingly likely to fail at their New Year's resolutions as the calendar year progresses. If flexible goals buffer against the negative consequences of goal violations, then those benefits should become more pronounced over time, thereby producing more durable benefits compared to inflexible goals.

Hypothesis 3: Subgoals that offer more flexibility (e.g., "every 2 weeks" goals) will produce longer lasting increases in goal progress than subgoals that offer less flexibility (e.g., "every week" goals).

We present the results of a preregistered field experiment to test these hypotheses. We also present a preregistered online vignette pilot study to confirm that we manipulated goal granularity and flexibility as intended and to test for evidence suggesting that the mechanisms we theorize could drive the benefits of subgoals might be at play.

Method

This research was approved by the institutional review board of the University of Pennsylvania (Protocol No. 831988 "Increasing Volunteer Motivation").

Transparency and Openness

We describe our sampling plan, data exclusions, manipulations, and measures in the study. We adhered to the *Journal of Applied Psychology* methodological checklist. Data, analysis code, and research materials are available at https://osf.io/7t8sz/?view_only=11668e16d2a44b8e972073472319dfca. Note that only a subset of our field experiment data can be made available due to their proprietary nature. Data were analyzed using R, Version 3.6.0 (R Core Team, 2019) and the packages aod, Version 1.3.1 (Lesnoff & Lancelot, 2012); fixest, Version 0.10.4 (Berge, 2018); lfe, Version 2.8-3 (Gaure, 2013); lmtest, Version 0.9-40 (Zeileis & Hothorn, 2002); MASS, Version 7.3-57 (Venables & Ripley, 2002); pscl, Version 1.5.2 (Zeileis et al., 2008); rio, Version 0.5.16 (Chan et al., 2021); rstatix, Version 0.7.0 (Kassambara, 2021); sandwich, Version 3.0-2 (Zeileis et al., 2020; Zeileis, 2006); and tidyverse, Version 1.2.1 (Wickham, 2017). The study design and analyses were preregistered at https://osf.io/fyhbx/files.

Field Experiment Setting

We conducted a preregistered field experiment in collaboration with Crisis Text Line (CTL), a national nonprofit organization that provides free crisis counseling via text message for a wide variety of issues, including suicidal ideation, mental health challenges, and abuse. Volunteers use an online text-messaging platform maintained by CTL—thus, volunteering takes place online and remotely and can be measured objectively by CTL.

At the time of our study, CTL had approximately 4,000 active volunteers in the United States, who all committed to an overarching goal of completing 200 hr of volunteering within a year of joining. Despite this target being oft-repeated to volunteers both during and after their volunteer training, as of October 31, 2019, fewer than 5% of volunteers who had been with the organization for at least a year had met their 200-hr commitment. While CTL does not penalize volunteers for failing to meet their 200-hr volunteering goal, this shortfall motivated the organization to explore interventions to boost volunteer motivation and goal pursuit.

Volunteers are encouraged to schedule weekly texting shifts, and CTL's online platform contains a scheduling tool that sends reminders prior to scheduled shifts.[3] The online platform also has a dashboard that allows volunteers to track how many hours they have volunteered. Whenever volunteers sign on to CTL's online platform, the landing page they see first displays the dashboard. In other words, volunteers see their dashboard every time they volunteer, and thus have easy access to feedback about goal progress.

Field Experiment Sample

CTL included all 3,805 active volunteers in the United States who had not yet volunteered 200 hr on February 4, 2019, in our experiment. In addition (following our preregistration), all new volunteers who joined CTL between February 5, 2019, and June 10, 2019, were also added to the experiment. This process led to a total sample size of $N = 9{,}108$ volunteers. Of the 34% ($n = 3{,}122$) of volunteers in our sample who opted to report their gender to CTL, 79% ($n = 2{,}477$) identified as female. Of the 99% ($n = 9{,}015$) who opted to report their age, the average age was 28.9 years (SD = 9.7 years).

Field Experiment Design

We randomly assigned all 9,108 volunteers in our study to three different experimental conditions: a maximally granular and inflexible subgoal condition, a moderately granular and flexible subgoal condition, and a maximally flexible and minimally granular control condition. In the maximally granular and inflexible subgoal condition ($n = 3{,}037$)—which we will refer to as the *4 hr every week* condition—volunteers were encouraged to reach their 200-hr volunteering goal by volunteering "4 hr every week." In the moderately granular and flexible subgoal condition ($n = 3{,}036$)—which we will refer to as the *8 hr every 2-week* condition—volunteers were encouraged to reach their 200-hr goal by volunteering "8 hr every 2 weeks" and were also told "for example, you can volunteer 6 hours in one week and 2 hours the next, or 4 hours every week" to ensure they understood they could reach this goal in different ways (i.e., flexibly). The maximally flexible and minimally granular control condition ($n = 3{,}035$)—which we will refer to as the *200 hr a year control* condition—which mirrored CTL's standard

[3] We were unable to obtain data from this scheduler tool. CTL estimated that roughly 50% of volunteers use the scheduler on a monthly basis (as of March 4, 2020).

messaging, volunteers were encouraged to reach their 200-hr goal by volunteering "some hours every week."

Our intervention was delivered through six emails, sent over the course of 12 weeks, all via CTL's usual email system for communicating with volunteers.[4] Each email offered a recommendation as to how volunteers should fulfill their 200-hr commitment: either by volunteering "some hours" every week, 4 hr every week, or 8 hr every 2 weeks (depending on volunteers' experimental condition). The first email also provided a visual example of a volunteering schedule displayed over a 2-week time interval, with four evenly distributed 2-hr shifts in the *4 hr every week* condition, and four unevenly distributed 2-hr shifts (three in the first week, one in the second week) in the *200 hr a year control* and *8 hr every 2-week* conditions. Finally, each email encouraged volunteers to schedule their volunteering hours for the next 2 weeks using CTL's online scheduling tool.

Volunteers then received five reminder emails, one every 2 weeks after the first email, to reinforce the initial message. These reminder emails reiterated the recommendations made in the original message. Complete email stimuli for the study can be found in online Supplemental Figures S1–S2.[5]

Pilot Study

Before presenting our analysis strategy and findings from this field experiment, we present results from a pilot vignette study. We ran this pilot study to confirm that our field experiment stimuli changed people's perceptions of goal granularity and flexibility as intended. We also ran this pilot to test for the mechanisms theorized to be at work in our field experiment.

Method

We recruited N = 900 participants on Mechanical Turk to complete a 5-min pilot survey for \$0.75 (44% identified as men; 77% identified as White; M_{age} = 41.7 years; SD_{age} = 13.5 years). This study was preregistered on AsPredicted.org (https://aspredicted.org/SYJ_4JC).

Pilot participants were asked to imagine they were volunteers at CTL—described as "a national nonprofit that provides text-based mental health support to people in need"—and to imagine that they had committed to volunteering 200 hr within a year. Participants were then asked to imagine that they received two versions of an email from CTL with different recommendations about how volunteers should work toward their 200-hr commitment.

Next, participants were shown abbreviated versions of two of the intervention emails sent in our field experiment, side by side. Participants were randomly assigned to either see the *200 hr a year control* condition and *4 hr every week* condition emails (labeled Emails 1 and 2 at random in counterbalanced order), the *200 hr a year control* condition and *8 hr every 2-week* condition emails (again with the same neutral, counterbalanced labels), or the *4 hr every week* condition and *8 hr every 2-week* condition emails (again with neutral and counterbalanced labels). Participants then answered a series of questions comparing the two emails on a 6-point Likert scale (reporting which email would elicit more of a certain reaction). These questions included two manipulation checks measuring the perceived (a) granularity and (b) flexibility of the goals described, as well as two-item measures of the degree to which they imagined receiving each email would (c) lead them to procrastinate on volunteering, adapted from Yockey (2016); Spearman–Brown = 0.90, (d) boost their self-efficacy, adapted from Giles et al. (2004); Spearman–Brown = 0.96, and (e) increase their goal commitment, adapted from Klein et al. (2001); Spearman–Brown = 0.95. Complete study materials are available in the online Supplemental Material.

Pilot Results

A correlation matrix of all variables collected in this study (Table S1) is available in the online Supplemental Material. Following our preregistered analysis plan, we ran one-sample t tests comparing each dependent variable within each comparison group to each Likert scale's midpoint (3.5), which would indicate no difference in which email would elicit more of a certain reaction.[6] First, our field study manipulations had the intended effect in our pilot. Participants reported that the *4 hr every week* email's goal was more granular, $M = 4.48$, $SD = 1.59$, $t(301) = 10.72$, $p < .001$, Cohen's $d = 0.617$, and less flexible, $M = 2.35$, $SD = 1.69$, $t(301) = -11.87$, $p < .001$, Cohen's $d = -0.683$, than the *200 hr a year control* email's goal. Participants also reported that the *8 hr every 2-week* email's goal was more granular, $M = 4.35$, $SD = 1.73$, $t(298) = 8.49$, $p < .001$, Cohen's $d = 0.491$, and less flexible, $M = 2.51$, $SD = 1.75$, $t(298) = -9.79$, $p < .001$, Cohen's $d = -0.566$, than the *200 hr a year control* email's goal. Finally, participants reported that the *8 hr every 2 weeks* email's goal was less granular, $M = 2.50$, $SD = 1.55$, $t(298) = -11.18$, $p < .001$, Cohen's $d = -0.647$, and more flexible, $M = 4.00$, $SD = 1.95$, $t(298) = 4.43$, $p < .001$, Cohen's $d = 0.256$, than the *4 hr every week* email's goal.

We also find evidence from our pilot that our manipulations elicit the theorized psychological processes. First, as theorized, pilot study participants forecasted that the *4 hr every week* email's goal would reduce procrastination, $M = 2.46$, $SD = 1.24$, $t(301) = -14.53$, $p < .001$, Cohen's $d = -0.836$, boost self-efficacy, $M = 4.40$, $SD = 1.56$, $t(301) = 10.07$, $p < .001$, Cohen's $d = 0.579$, and enhance goal commitment, $M = 4.61$, $SD = 1.43$, $t(301) = 13.43$, $p < .001$, Cohen's $d = 0.773$, relative to the *200 hr a year control* email. They also forecasted that the *8 hr every 2-week* email's goal would reduce procrastination, $M = 2.64$, $SD = 1.41$, $t(298) = -10.49$, $p < .001$, Cohen's $d = -0.607$, boost self-efficacy, $M = 4.09$, $SD = 1.77$, $t(298) = 5.77$, $p < .001$, Cohen's $d = 0.333$, and enhance goal commitment, $M = 4.34$, $SD = 1.66$, $t(298) = 8.78$, $p < .001$, Cohen's $d = 0.507$, relative to the *200 hr a year control* email. But compared to the *4 hr every week* email, participants forecasted that the *8 hr every 2-week* email's goal would produce more procrastination, $M = 4.05$, $SD = 1.29$, $t(298) = 7.39$, $p < .001$, Cohen's $d = 0.427$, less self-efficacy, $M = 2.94$, $SD = 1.66$, $t(298) = -5.78$, $p = .004$,

[4] Volunteers had the ability to opt out of receiving all emails from CTL at any point. Two hundred fifty-one participants (2.8% of our total sample) in our study availed themselves of this option during their intervention period and thus did not receive all six emails. Since our analyses are "intention to treat," this does not change who is included in our analyses or results.

[5] Our online Supplemental Material can be found here https://osf.io/dekjc?view_only=11668e16d2a44b8e972073472319dfca.

[6] We stated in our preregistration that we would standardize each scale item before analysis, which was an error given our analysis strategy. Our analysis relies on comparison against the unstandardized scale midpoint of 3.5, which means that standardizing scale items would invalidate this test. Therefore, we report results without standardizing any scale items.

Cohen's $d = -0.334$, and less goal commitment, $M = 2.80$, $SD = 1.50$, $t(298) = -8.03$, $p < .001$, Cohen's $d = -0.464$.

Consistent with our theorizing, more granular subgoals were predicted to reduce procrastination and boost self-efficacy and goal commitment. However, more granular subgoals were also seen as sacrificing flexibility. This comes with downsides based on our theorizing and prior research, such as a greater risk of goal abandonment in the face of setbacks, diminished performance on goals, and less control over one's schedule (Beshears et al., 2021; Moen et al., 2016; Sharif & Shu, 2017, 2021). Thus, this study supports our theorizing about the conflicting forces that come into play as goals become more granular but simultaneously less flexible.

Taken together, this pilot supports our theorizing regarding how our field experiment stimuli should change people's thinking about their goals.

Statistical Analyses of Field Experiment Data

Our field experiment's primary, preregistered dependent measure was the average number of minutes a participant volunteered for CTL each week during our study period. This was captured objectively by CTL as time spent on the organization's online volunteering platform. Following our preregistration, we log-transformed this dependent measure because we expected it to be skewed.[7] Our preregistered study period was divided into two phases. The first phase was a 12-week intervention period, during which participants received our intervention emails every 2 weeks. This phase started on the date when a participant received our first email and ended 2 weeks after they received our sixth and final reminder email.

The second phase of our study period was a 12-week postintervention period that immediately followed the intervention period. In this phase, participants no longer received any intervention emails. Following our preregistration, for each participant, we analyzed weekly data on their time spent volunteering during both our study's 12-week intervention period and 12-week postintervention period (i.e., a 24-week study period).[8]

To analyze our experimental data and assess the impact of our experimental treatments on participants' time spent volunteering during both the intervention period and the postintervention period, we relied on preregistered ordinary least squares (OLS) regressions analyses. We predicted each study participant's (log-transformed) weekly number of minutes volunteered during the preregistered 24-week study period. Our key independent variables were separate binary indicators for assignment to each of our subgoal treatment conditions (*4 hr every week* and *8 hr every 2 weeks*) during the relevant time period (the 12-week intervention period or the 12-week postintervention period) with an indicator omitted for the *200 hr a year control* condition. We included the following participant-level control variables in our primary preregistered analysis[9]: an indicator for whether a participant was male, an indicator for whether a participant's gender was unknown, a measure of a participant's age, an indicator for whether a participant's age was unknown, a tally of the total number of minutes a participant volunteered with CTL prior to their intervention start date, a tally of the total number of minutes a participant volunteered with CTL during the 4 weeks prior to their intervention start date, an indicator for whether a participant ever volunteered with CTL prior to their intervention start date, a measure of the number of days separating a participant's first-time volunteering for CTL and their intervention start date, and the number of weeks separating an observation and the corresponding intervention start date. We also included fixed effects for the calendar week of the year to account for seasonality. Finally, we clustered errors at the participant level to account for the longitudinal nature of our study. See online Supplemental Material for the exact specification of our OLS model.

In addition to obtaining data on the time participants spent volunteering for CTL during- and post-intervention, we also obtained weekly preintervention records of participants' time spent volunteering. We exploited these historical preintervention data in a second OLS regression specification, which was identical to our first model except that it replaced participant-level controls with participant fixed effects and included all available preintervention weeks of data on participants' hours spent volunteering.

Following the recommendation of Becker (2005), we present our findings both with control variables (as preregistered) and without control variables (to demonstrate robustness). The inclusion of controls does not change any of our key results, and we focus our discussion on our preregistered analyses, following best practices in open science (Logg & Dorison, 2021).

Results

Summary Statistics

Tables 1 and 2 summarize descriptive statistics from our study sample. Table 1 reports means for participant demographic variables and preintervention volunteering patterns, both for our full participant sample (Column 2) and for participants in each experimental condition (Columns 3–5). As shown in Column 6, *F* tests verify that randomization was balanced across conditions on observables before the intervention. Table 2 summarizes volunteering rates throughout our study period. Online Supplemental Table S2 presents a full correlation matrix of all variables analyzed and Tables S3–S4 summarize volunteering rates throughout our study period by experimental condition.

The skewness coefficient for weekly minutes volunteered during our 12-week intervention period was 4.05 (values further from 0 suggest greater skewness, while the sign of the coefficient indicates the direction of skewness), confirming that our outcome variable was positively skewed as expected (Joanes & Gill, 1998). Log-transforming this variable—following our preregistration—reduced the skewness coefficient to 1.14.

Goal Progress During the Intervention Period

Figure 2 shows the average weekly minutes volunteered by experimental condition during each week of the 12-week

[7] Specifically, our dependent variable is log(number of minutes volunteered +1) to handle cases where 0 min were volunteered, since log(0) is undefined.

[8] On May 28, 2019, we were informed of an error made by CTL in sending out emails to 1,611 participants, leading us to exclude these participants' data from the moment they received the incorrect email onwards (we retained their data prior to the incorrect email being sent). This process is detailed in an addendum to our preregistration (which can be found here https://osf.io/y9zsk/) as well as in our online Supplemental Material.

[9] See online Supplemental Material for an explanation for why we included these control variables, as well as a description of how each control variable was measured (Becker, 2005).

Table 1
Descriptive Statistics and Randomization Checks

Variable	Full sample	200 hr a year control condition	4 hr every week condition	8 hr every 2-week condition	*F* statistics
Proportion female	0.81 (0.39)	0.80 (0.40)	0.81 (0.39)	0.81 (0.39)	0.21 (p = .81)
Proportion not reporting gender	0.66 (0.48)	0.65 (0.48)	0.66 (0.48)	0.67 (0.47)	1.60 (p = .20)
Age (in years)	28.9 (9.7)	28.9 (9.8)	29.0 (9.7)	28.8 (9.6)	0.2 (p = .82)
Proportion not reporting age	0.01 (0.10)	0.01 (0.10)	0.01 (0.10)	0.01 (0.10)	0.13 (p = .88)
Proportion with no volunteering prior to intervention start date	0.09 (0.28)	0.08 (0.27)	0.10 (0.29)	0.09 (0.28)	2.10 s (p = .12)
Total volunteering before intervention start date (in hours)	29.5 (34.8)	28.9 (33.8)	29.4 (35.0)	30.2 (35.5)	1.01 (p = .36)
Volunteering during the 4 weeks prior to intervention start date (in hours)	3.73 (5.66)	3.66 (5.58)	3.72 (5.66)	3.81 (5.73)	0.58 (p = .56)
Number of months since first instance of volunteering	6.64 (8.94)	6.70 (8.95)	6.52 (8.85)	6.69 (9.01)	0.37 (p = .69)
N	9,108	3,035	3,037	3,036	

Note. This table reports means or proportions of key variables for the full sample and for each experimental condition separately. Standard deviations are reported in the first four columns in parentheses. The *F-statistics* column reports results from *F* tests as to whether the experimental conditions significantly differed on any key observable variables. Age was calculated by subtracting birth year from 2019, the year the intervention was run. The *proportion female* row uses data from the 3,122 participants who opted to report their gender, and the *age* (in years) row uses data from the 9,015 participants who opted to report their birth year. All other rows use data from the full sample of 9,108 participants.

intervention period and shows more volunteering in both the *4 hr every week* and *8 hr every 2-week* conditions than the *200 hr a year control* condition. Table 3 presents results from our primary regression models demonstrating that participants in the *4 hr every week* and *8 hr every 2-week* conditions volunteered significantly more during our 12-week intervention period than participants in the *200 hr a year control* condition (supporting Hypothesis 1). As Column 2 shows (using the model with participant-level control variables), compared to the *200 hr a year control* condition, our primary regression model estimates that assignment to the *4 hr every week* condition produced an 8.4% increase (corresponding to an average of 3.76 extra weekly minutes spent volunteering per person)[10] in weekly volunteering during our 12-week intervention period (p = 014; Cohen's d = 0.034), while assignment to the *8 hr every 2-week* condition produced a comparable 9.2% boost (corresponding to an average of 4.12 extra weekly minutes spent volunteering per person) in weekly volunteering (p = .008; Cohen's d = 0.053) over the same time period.[11] Similarly, Column 3 (which presents a model with participant fixed effects rather than a model including controls for various participant characteristics) estimates an 8.2% boost (corresponding to an average of 3.67 extra weekly minutes spent volunteering per person) in weekly volunteering in the *4 hr every week* condition (p = .021), and a 7.1% increase (corresponding to an average of 3.18 extra weekly minutes spent volunteering per person) in the *8 hr every 2-week* condition (p = .043), relative to the *200 hr a year control* condition. Wald tests comparing the estimated effects of our two treatment conditions show that these conditions did not significantly differ from one another in either model (ps > .78), which means we do not find support for Hypothesis 2.

Our outcome measure (weekly minutes spent volunteering) included many zeros (i.e., weeks where a participant did not volunteer at all), so we reanalyzed our intervention period data using a zero-inflated negative binomial model (ZINB) as an exploratory (i.e., not preregistered) test (see online Supplemental Table S5). A ZINB is a mixture model consisting of a binary logit model that predicts the excess zeros in the data and a negative binomial count model to predict the remaining count data.

We find in the binary model part of the ZINB (see Table S5, Column 1) that both subgoal treatments marginally reduced the likelihood of zero volunteering in a given week relative to the *200 hr a year control* condition ($b_{4hr} = -0.075$, $p_{4hr} = .075$, $OR = 0.93$; $b_{8hr} = -0.079$, $p_{8hr} = .059$, $OR = 0.92$), supporting Hypothesis 1. There was no significant difference between the two subgoal treatments (p = .914), however (not supporting Hypothesis 2). This means both subgoal treatments had a statistically indistinguishable impact on the decision to volunteer at all.

We find in the count model part of the ZINB (see Table S5, Column 2) that relative to the *200 hr a year control* condition, the *8 hr every 2-week* condition increased the number of minutes volunteered by 4.6% (corresponding to an average of 8.06 extra weekly minutes spent volunteering per person) after accounting for

[10] To estimate the extra weekly minutes spent volunteering per person, we calculated the average weekly minutes volunteered during the 12-week intervention in the *200 hr a year control* condition (44.8 min) and multiplied it by the corresponding percent change effect size.

[11] Since we are running OLS with a log-transformed dependent variable (y), our regression coefficients (β) correspond to percent changes in the dependent variable using the formula: $\%\Delta y = 100 \times (e^{\beta} - 1)$.

Table 2
Summary of Participant Volunteering During the Study Period

Time period	Proportion volunteering 4 hr or more	Minutes volunteered		
		10th percentile	*M* (*SD*)	90th percentile
Intervention period				
Week 1	12.6%	0	82.5 (142.2)	256
Week 2	8.9%	0	64.3 (127.1)	230
Week 3	8.2%	0	59.4 (119.0)	221
Week 4	6.7%	0	51.1 (111.8)	190
Week 5	5.9%	0	46.0 (102.6)	166
Week 6	5.2%	0	40.7 (101.4)	149
Week 7	4.5%	0	36.7 (93.0)	136
Week 8	4.1%	0	33.6 (88.0)	128
Week 9	4.0%	0	31.5 (86.7)	123
Week 10	3.5%	0	29.4 (85.6)	117
Week 11	3.0%	0	25.2 (76.7)	108
Week 12	3.0%	0	24.5 (79.4)	101
Postintervention period				
Week 13	3.5%	0	28.5 (83.6)	114.3
Week 14	3.0%	0	26.2 (81.5)	108
Week 15	3.4%	0	27.8 (85.8)	113
Week 16	2.7%	0	25.3 (79.9)	102
Week 17	3.3%	0	28.1 (88.9)	111
Week 18	2.8%	0	25.2 (84.2)	97
Week 19	2.7%	0	25.2 (89.6)	94
Week 20	2.4%	0	23.4 (80.6)	89
Week 21	2.6%	0	23.0 (85.1)	80
Week 22	2.6%	0	20.7 (74.4)	66.3
Week 23	3.0%	0	24.2 (82.7)	87.3
Week 24	2.8%	0	21.5 (81.4)	66

Note. This table reports summary statistics for the full sample ($N = 9{,}108$) both during the intervention period (Weeks 1–12) and immediately after the intervention period (Weeks 13–24).

the excess zeroes in the data ($b = 0.045$, $p = .038$), while the *4 hr every week* condition did not significantly increase minutes volunteered among those who made nonzero commitments of time ($b = 0.013$, $p = .549$), thus only providing partial support for Hypothesis 1.[12] Further, the *8 hr every 2-week* condition increased minutes volunteered marginally more than the *4 hr every week* condition by 3.3% (corresponding to an average of 5.86 extra weekly minutes spent volunteering per person)[13] providing some evidence for Hypothesis 2 ($b = 0.032$, $p = .063$). This means that after accounting for the decision of whether to volunteer at all, the *8 hr every 2-week* condition had a marginally larger, positive impact on the number of hours people volunteered than the *4 hr every week* condition.

Procrastination During the Intervention Period

We ran exploratory analyses to test whether our subgoal treatments successfully reduced procrastination as theorized. We first measured procrastination by looking at the number of days that passed in a given week before a participant first volunteered in that week (more days passed before a first volunteering session were interpreted as more procrastination). We ran a discrete-time survival analysis model with this outcome variable for each week in our intervention period using daily volunteering data (see online Supplemental Table S6, Column 1). We find that the *4 hr every week* condition significantly reduced procrastination on volunteering relative to the *200 hr a year control* condition ($b = 0.077$, $p = .035$, $OR = 1.08$), and the *8 hr every 2-week* condition also marginally significantly reduced procrastination on volunteering relative to the *200 hr a year control* condition ($b = 0.070$, $p = .056$, $OR = 1.07$). There was not a significant difference in procrastination between the two subgoal conditions ($p = .862$).

We also measured procrastination by examining the number of times that participants volunteered in a given week during the 12-week intervention period (more instances of volunteering in a week were interpreted as less procrastination). We ran an OLS regression model with this outcome variable (see online Supplemental Table S6, Column 2). We find that the *4 hr every week* condition significantly increased the number of times participants volunteered in a given week by 7.7% (from an average of 0.402 times per week to 0.433 times per week)[14] relative to the *200 hr a year control* condition ($b = 0.027$, $p = .033$). The *8 hr every 2-week* condition also increased the number of times participants volunteered in a given week by 10.2% (from an average of 0.402 times per week to 0.443 times per week) relative to the *200 hr a year control* condition ($b = 0.036$, $p = .006$). Again, however, there was not a significant difference between the two subgoal conditions ($p = .496$).

These results, combined with our pilot study data, suggest that, as theorized, breaking an overarching goal down into subgoals may have increased goal progress in part by reducing procrastination.

Treatment Effect Durability During the Intervention Period

To test Hypothesis 3, we also examine the durability of our treatment effects over the course of our 12-week intervention period. Following our preregistration, we ran a regression model where we interacted the indicators for each treatment condition with a continuous measure of the (mean-centered) week of the 12-week intervention period (see Table 4). We find that the effect of assignment to our *8 hr every 2-week* treatment does not significantly change over time relative to the *200 hr a year control* condition (Columns 2–3; $ps > .79$). However, the effect of assignment to our *4 hr every week* condition drops an estimated 0.8%–0.9% per week during the 12-week intervention period relative to the effect of assignment to our *200 hr a year control* condition. That decline is marginally significant in our regression specification with participant fixed effects ($p = .055$; see Column 3 of Table 4), but it is not statistically significant in our regression specification with participant-level control variables ($p = .109$; see Column 2 of Table 4).

To examine whether there were differences in treatment durability between the *8 hr every 2 weeks* and *4 hr every week* conditions, we ran Wald tests comparing the coefficient estimates on these interaction terms in Table 4, Columns 2–3. These interaction terms differed either

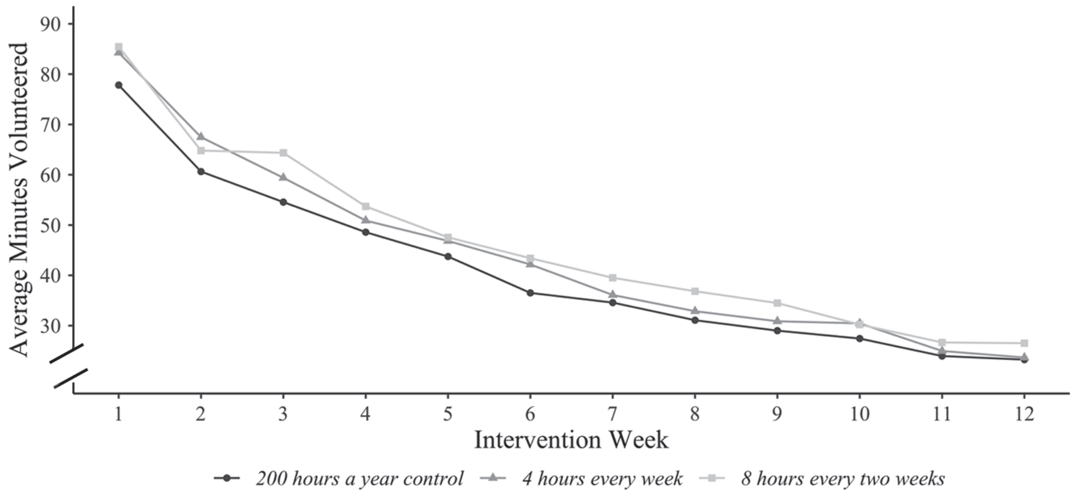
[12] Once again, we convert the coefficients in a negative binomial regression to percent changes in the dependent variable through exponentiation.

[13] To estimate the extra weekly minutes spent volunteering per person, we calculated the average weekly minutes volunteered—excluding instances of zero volunteering—during the 12-week intervention in the *200 hr a year control* condition (175.2 min) or the *4 hr every week* condition (177.5 min) and multiplied it by the corresponding percent change effect size. Note that this provides an imperfect estimated effect size, since the count model in the ZINB only accounts for excess zeros in the data, and not all zeros.

[14] To estimate the extra weekly times volunteered, we calculated the average weekly times volunteered during the 12-week intervention in the *200 hr a year control* condition (0.402 times) and multiplied it by the corresponding percent change effect size.

Figure 2

Average Weekly Minutes Volunteered During Each Week of the 12-Week Intervention Period



marginally in our preregistered OLS regression specification with controls (Column 2, p = .067) or significantly in our preregistered OLS regression specification with participant fixed effects (Column 3, p = .049). This provides suggestive evidence that the effect of the *8 hr every 2-week* treatment declines more slowly than the effect of the *4 hr every week* treatment, consistent with Hypothesis 3.

Goal Progress Postintervention

We also examine whether the effects of subgoal framing endured after our 12-week intervention concluded. Figure 3 displays the average weekly minutes volunteered by experimental condition during each week of the 12-week postintervention period and shows directionally more volunteering in both the *4 hr every week* and *8 hr every 2-week* conditions than the *200 hr a year control* condition. Table 3 presents the results of our preregistered regression models estimating the impact of our treatment conditions on volunteering in the postintervention period. As Columns 2–3 show, although estimates of the effects of assignment to the *4 hr every week* and *8 hr every 2-week* treatment conditions remained positive, they are not statistically significant postintervention (all ps > .22).

Table 3

Regression-Estimated Effects of Subgoal Treatments on Weekly Volunteering During- and Post-Iintervention

Independent variable	Dependent variable = weekly volunteering (log-transformed) (1)	(2)	(3)
During intervention			
4 hr every week	0.090* (0.042) p = .032	0.081* (0.033) p = .014	0.078* (0.034) p = .021
8 hr every 2 weeks	0.106* (0.042) p = .012	0.088** (0.033) p = .008	0.069* (0.034) p = .043
Postintervention			
4 hr every week	0.047 (0.036) p = .187	0.036 (0.032) p = .267	0.022 (0.038) p = .563
8 hr every 2 weeks	0.063+ (0.036) p = .080	0.039 (0.032) p = .226	0.014 (0.038) p = .720
Wald test comparing 4 hr versus 8 hr (during)	p = .708	p = .834	p = .784
Wald test comparing 4 hr versus 8 hr (post)	p = .671	p = .914	p = .830
Control variables?	No	Yes	No
Calendar week fixed effects?	No	Yes	Yes
Participant fixed effects?	No	No	Yes
Preintervention data?	No	No	Yes
Number of participants	9,108	9,108	9,108
Observations	189,784	189,784	485,082
R^2	0.022	0.203	0.317

Note. OLS = ordinary least squares. This table shows results from OLS regressions predicting weekly volunteering in number of minutes (log-transformed) across the experimental conditions for the 12-week intervention period, as well as the 12-week postintervention period. Robust standard errors, clustered at the participant level, are in parentheses. The *control* condition, as described in text, serves as the reference category. When controls are present, regressions include gender, age, total number of minutes volunteered prior to the intervention start date, number of minutes volunteered during the 4 weeks prior to the intervention start date, tenure as a volunteer, and the number of weeks since the intervention start date.

$^{+}p < .1$. $^{*}p < .05$. $^{**}p < .01$.

Table 4
Regression-Estimated Effects of How Subgoal Treatment Effects Change Over Time During 12-Week Intervention Period

	Dependent variable = log-transformed number of minutes volunteered in a given week		
Independent variable	(1)	(2)	(3)
4 hr every week	0.090* (0.042) p = .032	0.080* (0.033) p = .013	0.126** (0.039) p = .001
8 hr every 2 weeks	0.106 (0.042) p = .012	0.086** (0.033) p = .008	0.069+ (0.039) p = .079
Weeks since start of intervention (centered)	−0.096*** (0.004) p < .001	−0.112*** (0.007) p < .001	−0.113*** (0.004) p < .001
4 hr every week × Weeks since start of intervention	−0.007 (0.005) p = .162	−0.008 (0.005) p = .109	−0.009+ (0.005) p = .055
8 hr every 2 weeks × Weeks since start of intervention	0.002 (0.005) p = .646	0.001 (0.005) p = .800	0.0004 (0.005) p = .938
Wald test comparing 4-hr versus 8-hr interaction terms	p = .066+	p = .067+	p = .049*
Control variables?	No	Yes	No
Calendar week fixed effects?	No	Yes	Yes
Participant fixed effects?	No	No	Yes
Preintervention data?	No	No	Yes
Number of participants	9,108	9,108	9,108
Observations	99,808	99,808	395,106
R^2	0.023	0.240	0.318

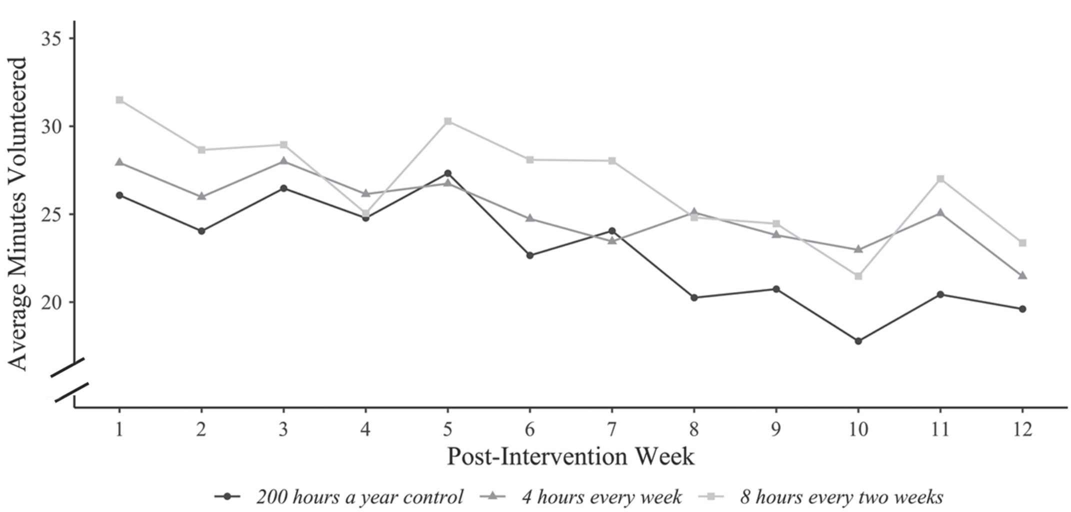
Note. OLS = ordinary least squares. This table presents a series of OLS regressions predicting weekly volunteering in number of minutes (log-transformed) across the experimental conditions for the 12-week intervention period. *Weeks since start of intervention* ranged from 1 to 12 during the intervention period, takes the value 0 in preintervention data, and was mean-centered. Robust standard errors, clustered at the participant level, are in parentheses. When controls are present, regressions include gender, age, total number of minutes volunteered prior to the intervention start date, number of minutes volunteered during the 4 weeks prior to the intervention start date, and tenure as a volunteer.
+ p < .1. * p < .05. ** p < .01. *** p < .001.

We also do not see differences between the decay rates of our two treatment conditions in the 12-week postintervention period (see online Supplemental Table S7).

Robustness Checks

We performed several robustness checks, all of which yielded results largely consistent with those generated by our preregistered analyses, and all of which can be found in the online Supplemental Material (Tables S5 and S8–S14). First, we analyzed the raw number of weekly minutes volunteered as the dependent variable (instead of analyzing the log of this variable to account for skew). Second, we analyzed a binary dependent variable capturing whether or not participants volunteered at all in a given week, and we analyzed this outcome with both OLS and logistic regressions. Third, we analyzed a binary dependent variable capturing whether or not participants volunteered at least 4 hr in a given week (which would keep them on track for 200 hr of yearly volunteering), and again, we analyzed this outcome with both OLS and logistic regressions. Fourth, we reran our regression models analyzing 2 weeks of volunteering at a time rather than week-by-week volunteering to align with the cadence of our emails to participants

Figure 3
Average Weekly Minutes Volunteered During Each Week of the 12-Week Postintervention Period



(which were sent every 2 weeks). Fifth, we reran our primary regression model first predicting the total number of minutes participants volunteered during the entire study and then predicting the total number of minutes volunteered during the entire poststudy period. Sixth, we reran our analyses using repeated-measures analysis of variance (ANOVA) and analysis of covariance (ANCOVA) instead of OLS. Finally, we tested the robustness of our ZINB results by running a two-stage OLS model. These analyses all support the same conclusions presented previously.

Exploration of possible heterogeneity in treatment effects revealed a lack of variation by volunteers' gender, tenure, or prior progress toward their 200-hr goal (see online Supplemental Tables S15–S19). However, treatment effects did vary by volunteers' age, such that older volunteers' volunteering time responded more positively to the *8 hr every 2 weeks* framing than either the *200 hr a year control* framing or the *4 hr every week* framing (see online Supplemental Table S20 and section "Additional Details on Heterogeneity by Volunteer Age").

General Discussion

Previous scholarship examining how subgoals affect goal progress has largely relied on one-shot studies in the laboratory (Amir & Ariely, 2008; Latham & Brown, 2006; Latham & Seijts, 1999; Stock & Cervone, 1990) or small-sample studies in the field (Bandura & Schunk, 1981; Bandura & Simon, 1977; Huang et al., 2017; Latham & Brown, 2006) to measure impact. In a large, preregistered, longitudinal field experiment, we show that subgoals substantially and robustly increase goal progress. Specifically, we find that breaking down volunteers' annual 200-hr commitment to a nonprofit into more granular subgoals—to volunteer either 4 hr every week (a granular and relatively inflexible subgoal) or 8 hr every 2 weeks (a less granular and more flexible subgoal)—boosts minutes volunteered by 7%–8% over a period of several months. The benefits of subgoal framing do not vary by gender, tenure, or prior goal progress. In addition, our preregistered vignette pilot study provides suggestive evidence that granular subgoals may boost goal progress by improving self-efficacy and goal commitment and by reducing procrastination.

However, there is a trade-off between granularity and flexibility in goals because breaking overarching goals into more granular subgoals also involves reducing the amount of flexibility available to goal seekers. Despite the greater flexibility afforded by describing volunteers' subgoal as contributing "8 hr every 2 weeks" rather than "4 hr every week" and the demonstrated benefits of goal flexibility (Beshears et al., 2021; Sharif & Shu, 2017, 2021), we largely do not find significant productivity differences between the two subgoal conditions as hypothesized. That said, we do find suggestive evidence in an exploratory ZINB analysis that conditional on volunteering more than 0 min in a week, the more flexible subgoal does boost people's time spent volunteering relative to the more inflexible subgoal. Moreover, we find that this more flexible and less granular subgoal also produces more durable benefits than the more granular, less flexible "4 hr every week" subgoal, as hypothesized. Together, these results suggest that goal flexibility may matter more when it comes to enhancing and maintaining goal commitment over time than motivating initial goal pursuit. Future research replicating these patterns and delving further into the underlying mechanisms would be valuable.

Theoretical Contributions

Our work contributes to the goal-setting literature in several ways. First, in a large, preregistered, longitudinal experiment, we demonstrate the robustness of the theory that subgoals reliably increase goal progress. Some have theorized and provided data from the laboratory suggesting that subgoals can fuel complacency and impair goal striving in environments where progress toward goals is observable (Amir & Ariely, 2008). However, our field data suggest that subgoals create value even under these circumstances. Building on past theorizing put forth by Gal and McShane (2012), we propose that this may be because, over longer time horizons (e.g., months rather than minutes), people's attention refocuses on the overarching goal, which reduces the potential risks of subgoals (e.g., complacency).

We also extend past theory by positing benefits of subgoals that have previously been overlooked. One potential benefit of subgoals we propose is that they lead to the creation of more imminent deadlines, which have been shown to effectively combat procrastination (Ariely & Wertenbroch, 2002; Janakiraman & Ordóñez, 2012; Lieberman et al., 2021). We test for this in our field experiment by looking at the number of days that passed before a participant first volunteered in a given week, and we find evidence that both of our subgoal treatments reduced this form of procrastination. We also find that both subgoal treatments increased the number of times participants volunteered in a given week, which we interpret as additional evidence of reduced procrastination.

Another possible benefit of subgoals that has been overlooked is that by design, they require smaller time commitments than overarching goals. This should make goal commitment more attractive, similar to the "pennies-a-day" effect (Gourville, 1998; Hershfield et al., 2020). We hope our investigation spurs future research on these and other benefits of subgoal framing.

Finally, we expand the goal-setting literature by exploring how the degree of flexibility in a subgoal may affect its impact, both theoretically and in practice. Prior research has operationalized goal flexibility in many different ways. Depending on the way it is operationalized, flexibility has been found to sometimes be helpful for goal pursuit (Beshears et al., 2021; Sharif & Shu, 2017), sometimes harmful (Koch & Nafziger, 2020; Shin & Milkman, 2016), and sometimes have no effect (Scott & Nowlis, 2013; van Lent, 2019). We manipulate goal flexibility by reframing the same overarching goal in a more or less granular way rather than by objectively limiting flexibility. We find evidence that with increasing granularity and reduced flexibility comes a cost in terms of slightly steeper declines in goal progress over time.

Limitations and Future Research

Our research has several important limitations. First, our sample came from a single U.S. organization's volunteer workforce. Additional research replicating and extending these findings to other samples, contexts, outcomes, and time periods would be valuable (List, 2020). For example, future work is necessary to establish whether our results would generalize to companies that pay their employees rather than those that rely on volunteers. Furthermore, the rates at which volunteers actually attained their overall 200-hr volunteering goal (or their subgoal of volunteering 4 hr every week or 8 hr every 2 weeks) were very low in our field experiment.

It would be valuable for future work to examine whether interventions like the one we tested might produce different results in settings where average rates of goal attainment are higher, and where the goals under study might therefore be perceived as more feasible.

Second, our intervention lasted only 12 weeks, and as is typical in the literature, after our reminders reframing goals concluded, their benefits dissipated (Calzolari & Nardotto, 2017). It would be valuable for future research to test how long the benefits of an intervention reframing overarching goals into more granular subgoals can endure if reminders continue to be delivered not for months but for years. It would also be valuable to explore whether educating people about how to break overarching goals into subgoals can produce lasting benefits that persist in the absence of reminders (Hertwig & Grüne-Yanoff, 2017). While our field experiment focused on an organizational application of subgoal framing, future work should also explore how this strategy can be used to motivate individual goal pursuit. For example, if individuals are pursuing large goals in the workplace, it would likely be beneficial for them to break those large goals down into more granular subgoals. Future work should also explore whether this strategy can be used in the domain of personal goal pursuit, for example, if individuals can learn to successfully create subgoals to achieve their personal ambitions, such as practicing a new skill, exercising, or healthy eating.

We also test the effects of subgoals in the context of a simple, albeit long-term, unidimensional goal. While there are many organizational settings involving unidimensional goals like this (e.g., in sales, call centers, or other volunteering contexts), it is worth considering what may happen in the context of more complex goals. In some settings, people work toward multifaceted goals that can be achieved by pursuing different types of subgoals (e.g., an overarching goal to "become a better software engineer" can be pursued by achieving different subgoals like "learning new programming languages," "doing more code reviews," "practicing important algorithms"). In these settings, past research suggests that subgoals may distract people from their overarching goals (Fishbach & Dhar, 2005; Fishbach et al., 2006). In addition, these contexts may raise the risk of goal substitution, whereby progress on one type of subgoal distracts people from making progress on other types of subgoals, ultimately harming progress toward their overarching goal (Fishbach & Dhar, 2005; Fishbach et al., 2006). There are also cases where people are simultaneously pursuing multiple distinct goals (e.g., concurrent goals to "read more research papers" and "engage in more service work"). Future research is needed to explore how subgoal framing at different levels of granularity and flexibility would apply in these different types of goal contexts.

Finally, we proposed several novel mechanisms to explain the benefits of subgoals in our field experiment, and our pilot study provides supportive evidence for these mechanisms, as does our exploratory analysis of how long people procrastinated each week in our field data. However, future work is needed to more directly establish which mechanisms drive the benefits of subgoals in field contexts like the one we studied.

Practical Implications

At an individual level, the gains from reframing an overarching volunteering goal as a series of more granular weekly or biweekly subgoals are on the order of magnitude of a few extra minutes volunteered every week. However, when scaled across a large organization like CTL over time, the 8% lift in volunteering we generate is much more meaningful. For example, if CTL rolled out our best-performing treatment across all of its volunteers for a year's time, we could expect it to produce an estimated 19,900 hr of additional volunteering (bootstrapped 95% confidence interval: 14,200 hr to 25,500 hr) at essentially zero cost.[15]

Thus, our findings suggest that subgoal framing can be a cost-effective and powerful tool for managers and organizations to motivate goal progress over time. There may also be ways to amplify the benefits of reframing an overarching goal into more granular subgoals by communicating via more channels than emails sent every other week.

[15] Calculations based on a mean difference of 5.75 min in weekly minutes volunteered between the *200 hr a year control* condition and the *8 hr every 2-week* condition from the data, scaled across 4,000 volunteers over a 52-week period. The confidence interval was calculated by (a) resampling within the *200 hr a year control* condition, (b) resampling within the *8 hr every 2-week* condition, (c) computing the mean difference between the two resampled groups, (d) repeating steps (a)–(c) 10,000 times, and (e) taking the 2.5th and 97.5th percentiles of the mean differences.

References

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin*, *82*(4), 463–496. https://doi.org/10.1037/h0076860

Amar, M., Ariely, D., Ayal, S., Cryder, C. E., & Rick, S. I. (2011). Winning the battle but losing the war: The psychology of debt management. *Journal of Marketing Research*, *48*(special issue), S38–S50. https://doi.org/10.1509/jmkr.48.SPL.S38

Amir, O., & Ariely, D. (2008). Resting on laurels: The effects of discrete progress markers as subgoals on task performance and preferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1158–1171. https://doi.org/10.1037/a0012857

Ariely, D., & Wertenbroch, K. (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science*, *13*(3), 219–224. https://doi.org/10.1111/1467-9280.00441

Bandura, A., & Cervone, D. (1986). Differential engagement of self-reactive influences in cognitive motivation. *Organizational Behavior and Human Decision Processes*, *38*(1), 92–113. https://doi.org/10.1016/0749-5978(86)90028-2

Bandura, A., & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology*, *88*(1), 87–99. https://doi.org/10.1037/0021-9010.88.1.87

Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology*, *41*(3), 586–598. https://doi.org/10.1037/0022-3514.41.3.586

Bandura, A., & Simon, K. M. (1977). The role of proximal intentions in self-regulation of refractory behavior. *Cognitive Therapy and Research*, *1*(3), 177–193. https://doi.org/10.1007/BF01186792

Becker, T. E. (2005). Potential problems in the statistical control of variables in organizational research: A qualitative analysis with recommendations. *Organizational Research Methods*, *8*(3), 274–289. https://doi.org/10.1177/1094428105278021

Berge, L. (2018). *Efficient estimation of maximum likelihood models with multiple fixed-effects: The R package FENmlm*. CREA Discussion Papers.

Beshears, J., Lee, H. N., Milkman, K. L., Mislavsky, R., & Wisdom, J. (2021). Creating exercise habits using incentives: The trade-off between flexibility and routinization. *Management Science*, *67*(7), 3985–4642. https://doi.org/10.1287/mnsc.2020.3706

Calzolari, G., & Nardotto, M. (2017). Effective reminders. *Management Science*, *63*(9), 2915–2932. https://doi.org/10.1287/mnsc.2016.2499

Chan, C., Chan, G. C., Leeper, T. J., & Becker, J. (2021). *rio: A Swiss-army knife for data file I/O* (R package Version 0.5.29) [Computer software].

Cochran, W., & Tesser, A. (1996). The "What the Hell" effect: Some effects of goal proximity and goal framing on performance. In L. L. Martin & A. Tesser (Eds.), *Striving and feeling: Interactions among goals, affect, and self-regulation* (pp. 99–120). Lawrence Erlbaum Associates. https://psycnet.apa.org/record/1996-97873-004

Fishbach, A., & Dhar, R. (2005). Goals as excuses or guides: The liberating effect of perceived goal progress on choice. *The Journal of Consumer Research*, *32*(3), 370–377. https://doi.org/10.1086/497548

Fishbach, A., Dhar, R., & Zhang, Y. (2006). Subgoals as substitutes or complements: The role of goal accessibility. *Journal of Personality and Social Psychology*, *91*(2), 232–242. https://doi.org/10.1037/0022-3514.91.2.232

Frese, M., & Zapf, D. (1994). Action as the core of work psychology: A German approach. In H. C. Triandis, M. D. Dunnette, & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 4, pp. 271–340). Consulting Psychologists Press.

Gal, D., & McShane, B. B. (2012). Can small victories help win the war? Evidence from consumer debt management. *JMR, Journal of Marketing Research*, *49*(4), 487–501. https://doi.org/10.1509/jmr.11.0272

Gaure, S. (2013). OLS with multiple high dimensional category variables. *Computational Statistics & Data Analysis*, *66*, 8–18. https://doi.org/10.1016/j.csda.2013.03.024

Giles, M., McClenahan, C., Cairns, E., & Mallet, J. (2004). An application of the Theory of Planned Behaviour to blood donation: The importance of self-efficacy. *Health Education Research*, *19*(4), 380–391. https://doi.org/10.1093/her/cyg063

Gourville, J. T. (1998). Pennies-a-day: The effect of temporal reframing on transaction evaluation. *The Journal of Consumer Research*, *24*(4), 395–408. https://doi.org/10.1086/209517

Heath, C., Larrick, R. P., & Wu, G. (1999). Goals as reference points. *Cognitive Psychology*, *38*(1), 79–109. https://doi.org/10.1006/cogp.1998.0708

Herman, C. P., & Mack, D. (1975). Restrained and unrestrained eating. *Journal of Personality*, *43*(4), 647–660. https://doi.org/10.1111/j.1467-6494.1975.tb00727.x

Hershfield, H. E., Shu, S., & Benartzi, S. (2020). Temporal reframing and participation in a savings program: A field experiment. *Marketing Science*, *39*(6), 1039–1051. https://doi.org/10.1287/mksc.2019.1177

Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, *12*(6), 973–986. https://doi.org/10.1177/1745691617702496

Huang, S. C., Jin, L., & Zhang, Y. (2017). Step by step: Sub-goals as a source of motivation. *Organizational Behavior and Human Decision Processes*, *141*, 1–15. https://doi.org/10.1016/j.obhdp.2017.05.001

Janakiraman, N., & Ordóñez, L. (2012). Effect of effort and deadlines on consumer product returns. *Journal of Consumer Psychology*, *22*(2), 260–271. https://doi.org/10.1016/j.jcps.2011.05.002

Joanes, D. N., & Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *Statistician*, *47*(1), 183–189. https://doi.org/10.1111/1467-9884.00122

Kassambara, A. (2021). *rstatix: Pipe-friendly framework for basic statistical tests* (R package Version 0.7.0) [Computer software]. https://CRAN.R-project.org/package=rstatix

Kelly, E. L., Moen, P., & Tranby, E. (2011). Changing workplaces to reduce work-family conflict: Schedule control in a white-collar organization. *American Sociological Review*, *76*(2), 265–290. https://doi.org/10.1177/0003122411400056

Klein, H. J., Wesson, M. J., Hollenbeck, J. R., Wright, P. M., & DeShon, R. P. (2001). The assessment of goal commitment: A measurement model meta-analysis. *Organizational Behavior and Human Decision Processes*, *85*(1), 32–55. https://doi.org/10.1006/obhd.2000.2931

Koch, A. K., & Nafziger, J. (2020). Motivational goal bracketing: An experiment. *Journal of Economic Theory*, *185*, Article 104949. https://doi.org/10.1016/j.jet.2019.104949

Krause, K., & Freund, A. M. (2014). How to beat procrastination. *European Psychologist*, *19*(2), 132–144. https://doi.org/10.1027/1016-9040/a000153

Latham, G. P., & Brown, T. C. (2006). The effect of learning vs. outcome goals on self-efficacy, satisfaction and performance in an MBA program. *Applied Psychology*, *55*(4), 606–623. https://doi.org/10.1111/j.1464-0597.2006.00246.x

Latham, G. P., & Seijts, G. H. (1999). The effects of proximal and distal goals on performance on a moderately complex task. *Journal of Organizational Behavior*, *20*(4), 421–429. https://doi.org/10.1002/(SICI)1099-1379(199907)20:4<421::AID-JOB896>3.0.CO;2-%23

Lesnoff, M., & Lancelot, R. (2012). *aod: Analysis of overdispersed data* (R package Version 1.3.2) [Computer software]. https://cran.r-project.org/package=aod

Lieberman, A., Gneezy, A., Berry, E., Miller, S., Koch, M., Argenbright, K. E., & Gupta, S. (2021). The effect of deadlines on cancer screening completion: A randomized controlled trial. *Scientific Reports*, *11*(1), Article 13876. https://doi.org/10.1038/s41598-021-93334-1

List, J. A. (2020). *Non est disputandum de generalizability? A glimpse into the external validity trial*. NBER Working Paper (w27535).

Locke, E. A., & Latham, G. P. (2019). The development of goal setting theory: A half century retrospective. *Motivation Science*, *5*(2), 93–105. https://doi.org/10.1037/mot0000127

Logg, J. M., & Dorison, C. A. (2021). Pre-registration: Weighing costs and benefits for researchers. *Organizational Behavior and Human Decision Processes*, *167*, 18–27. https://doi.org/10.1016/j.obhdp.2021.05.006

Moen, P., Kelly, E. L., Fan, W., Lee, S. R., Almeida, D., Kossek, E. E., & Buxton, O. M. (2016). Does a flexibility/support organizational initiative improve high-tech employees' well-being? Evidence from the work, family, and health network. *American Sociological Review*, *81*(1), 134–164. https://doi.org/10.1177/0003122415622391

Norcross, J. C., & Vangarelli, D. J. (1988–1989). The resolution solution: Longitudinal examination of New Year's change attempts. *Journal of Substance Abuse*, *1*(2), 127–134. https://doi.org/10.1016/S0899-3289(88)80016-6

Oscarsson, M., Carlbring, P., Andersson, G., & Rozental, A. (2020). A large-scale experiment on New Year's resolutions: Approach-oriented goals are more successful than avoidance-oriented goals. *PLOS ONE*, *15*(12), Article e0234097. https://doi.org/10.1371/journal.pone.0234097

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Schouwenburg, H. C., & Groenewoud, J. (2001). Study motivation under social temptation; effects of trait procrastination. *Personality and Individual Differences*, *30*(2), 229–240. https://doi.org/10.1016/S0191-8869(00)00034-9

Scott, M. L., & Nowlis, S. M. (2013). The effect of goal specificity on consumer goal reengagement. *The Journal of Consumer Research*, *40*(3), 444–459. https://doi.org/10.1086/670766

Seijts, G. H., & Latham, G. P. (2001). The effect of distal learning, outcome, and proximal goals on a moderately complex task. *Journal of Organizational Behavior*, *22*(3), 291–307. https://doi.org/10.1002/job.70

Sharif, M. A., & Shu, S. B. (2017). The benefits of emergency reserves: Greater preference and persistence for goals that have slack with a cost. *JMR, Journal of Marketing Research*, *54*(3), 495–509. https://doi.org/10.1509/jmr.15.0231

Sharif, M. A., & Shu, S. B. (2021). Nudging persistence after failure through emergency reserves. *Organizational Behavior and Human Decision Processes*, *163*, 17–29. https://doi.org/10.1016/j.obhdp.2019.01.004

Shin, J., & Milkman, K. L. (2016). How backup plans can harm goal pursuit: The unexpected downside of being prepared for failure. *Organizational Behavior and Human Decision Processes*, *135*, 1–9. https://doi.org/10.1016/j.obhdp.2016.04.003

Soman, D., & Cheema, A. (2004). When goals are counterproductive: The effects of violation of a behavioral goal on subsequent performance. *The Journal of Consumer Research*, *31*(1), 52–62. https://doi.org/10.1086/383423

Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin*, *133*(1), 65–94. https://doi.org/10.1037/0033-2909.133.1.65

Stock, J., & Cervone, D. (1990). Proximal goal-setting and self-regulatory processes. *Cognitive Therapy and Research*, *14*(5), 483–498. https://doi.org/10.1007/BF01172969

van Lent, M. (2019). Goal setting, information, and goal revision: A field experiment. *German Economic Review*, *20*(4), e949–e972. https://doi.org/10.1111/geer.12199

Vancouver, J. B., Thompson, C. M., Tischner, E. C., & Putka, D. J. (2002). Two studies examining the negative effect of self-efficacy on performance. *Journal of Applied Psychology*, *87*(3), 506–516. https://doi.org/10.1037/0021-9010.87.3.506

Vancouver, J. B., Thompson, C. M., & Williams, A. A. (2001). The changing signs in the relationships among self-efficacy, personal goals, and performance. *Journal of Applied Psychology*, *86*(4), 605–620. https://doi.org/10.1037/0021-9010.86.4.605

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer. https://doi.org/10.1007/978-0-387-21706-2

Wickham, H. (2017). *tidyverse: Easily install and load the 'Tidyverse'*. R package Version 1.2.1. https://CRAN.R-project.org/package=tidyverse

Yockey, R. D. (2016). Validation of the short form of the academic procrastination scale. *Psychological Reports*, *118*(1), 171–179. https://doi.org/10.1177/0033294115626825

Zeileis, A. (2006). Object-oriented computation of sandwich estimators. *Journal of Statistical Software*, *16*(9), 1–16. https://doi.org/10.18637/jss.v016.i09

Zeileis, A., & Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, *2*(3), 7–10. https://CRAN.R-project.org/doc/Rnews/

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, *27*(8), 1–25. https://doi.org/10.18637/jss.v027.i08

Zeileis, A., Köll, S., & Graham, N. (2020). Various versatile variances: An object-oriented implementation of clustered covariances in R. *Journal of Statistical Software*, *95*(1), 1–36. https://doi.org/10.18637/jss.v095.i01

Zhu, M., Bagchi, R., & Hock, S. J. (2019). The mere deadline effect: Why more time might sabotage goal pursuit. *The Journal of Consumer Research*, *45*(5), 1068–1084. https://doi.org/10.1093/jcr/ucy030

Received October 27, 2021
Revision received June 29, 2022
Accepted June 30, 2022 ■