Targeting Behavioral Change Interventions: An Experiment on Debiasing Savings for College *

Guglielmo Briscese † Sabrina Liu‡
Job Market Paper

October 14, 2024
Latest version here.

Interventions to change behaviors that are detrimental to decision-makers may produce different outcomes for different individuals. Previous research has treated these interventions as both diagnostic and solution tools, limiting insights into the sources of treatment effect heterogeneity and their replicability across contexts. Our study addresses this gap by implementing a theory-driven experiment: we invite 3,825 parents with college savings accounts to participate in a short survey designed to (i) identify and quantify their behavioral biases, and (ii) administer a message to correct these biases. We find that behavioral biases are common, not systematically linked to observable characteristics, and correlated with lower savings. By tracking savings balance for one year, we show that messages that address an individual's specific biases are more effective in increasing savings than they are on individuals who display different biases. We provide a framework for assessing the scalability and cost-effectiveness of similar targeted behavioral change interventions. Our results highlight the need for distinguishing between diagnostic and solution tools to generate insights on the sources of treatment effect heterogeneity and intervention replicability across contexts.

JEL Classification Codes: D03, D91, H43, I38.

Keywords: Personalized nudges; RCTs; College Savings Accounts.

*This study obtained IRB approval from University of Chicago: IRB21-1268. We thank John Mitchell and the staff at the Illinois State Treasurer's Office staff for their collaboration, Jesse Orr for impeccable project management, Noah Sebek and Saatvik Amravathi for excellent research assistance, and Pietro Ramella for insightful discussions. We thank for useful comments: John List, Susan Athey, Abigail Sussman, Mario Macis, Eugenio Proto, Giulio Zanella, Nicola Lacetera, Kelly Hallberg, attendees at the University of Chicago Booth Behavioral Economics Lab and Department of Economics Experimental seminars, as well as participants at the International Behavioural Public Policy Conference and the Behavioral Science and Policy Association Conference.

†Corresponding author, Post-doctoral Scholar, The University of Chicago, gubri@uchicago.edu.
‡Lead analyst, The University of Chicago, liusabrina@uchicago.edu.

1 Introduction

Governments and private sector organizations around the world have been increasingly adopting behavioral change interventions to improve outcomes across various policy areas (OECD and Policy, 2017). These interventions, often evaluated using randomized control trials, aim to influence behavior through cost-effective and scalable solutions. However, their efficacy has been inconsistent, creating uncertainty for researchers and policymakers about their replicability in different contexts. This study posits that previous studies may suffer from a *hypothesis bending* limitation: by using these interventions as both diagnostic and solution tools, these studies interpret positive treatment effects as evidence of both the behavioral barrier's existence and the solution's efficacy. This limits our understanding of why individuals behave suboptimally and why some interventions work better than others or might not replicate elsewhere.

We adopt a different approach by first identifying why individuals behave suboptimally and then running a randomized intervention to see if its efficacy depends on its relevance to the individual's specific behavioral bias, whether this being a detrimental habit or a biased belief, that the person experiences. Failing to account for this important but largely overlooked source of heterogeneity can lead to mis-estimation of treatment effects and low replicability.

We partnered with a government agency to conduct a theory-driven field experiment aimed at encouraging greater savings - purposely choosing a well-studied policy area in this literature. The goal is to encourage parents to increase contributions to their children's college savings accounts via light-touch scalabe digital messages. We ran a meta-analysis of previous studies testing similar interventions to increase savings to avoid duplicating ineffective solutions and found evidence of publication bias and inconsistent results, highlighting the need for a new experiment. We thus invited 3,825 parents with college savings accounts to participate in a paid online survey, eliciting their beliefs, habits, and preferences to identify potential behavioral barriers hindering their savings. Survey respondents were then randomly assigned to different final survey pages with various nudge messages aimed at encouraging

greater savings, resembling behaviorally-informed interventions tested in previous studies (see Appendix E in Section 12.4 for a set of related studies). We matched the survey data with historical administrative records and tracked their savings balances for one year.

We find, first, that the behavioral barriers we identified are common among this group despite sharing similar saving goals and socio-economic characteristics. Common suboptimal behavioral patterns include the reliance on automatic transfers, which we show may anchor people to lower savings in the long run; the belief that one's saving amount is comparable to their peers; an underestimation of the cost of college by not accounting for (low) inflation of college cost by the time children reaches college age; and the underestimation of the value of small contributions by not considering the power of compound interest. Second, these barriers are linked to significantly lower savings, regardless of the income of the account owner or the lifespan of a savings account. After controlling for all available characteristics, these barriers are correlated with an average of $14,000 less in savings balances. Third, the relevance of the nudge matters: respondents who received messages addressing their specific behavioral barriers were significantly more likely to make an extra deposit within 30 days of the survey. One year later, this effect translated into an additional $50 saved per month (from a baseline of $462) compared to parents who were in the control group as well as those who saw the same messages but for whom they were irrelevant. Using a simple theoretical model, we identify the conditions under which diagnosing behavioral barriers is valuable for social planners seeking to implement targeted nudges. We also provide ideas for replication and scalability, and demonstrate how our trial implementation partner successfully scaled this approach across the state.

Overall, our study underscores the need for a deeper understanding of human behavior before deploying large-scale interventions. Since behaviors are context-dependent, investing in greater personalization can be more cost-effective. This approach highlights the importance of contextual expertise prior to the design and implementation of behavioral change interventions, acknowledging the complexity and variability of human behavior.

Related literature and contribution. Our study contributes to multiple streams

of the literature. We address the growing demand within the social sciences for a better understanding of heterogeneous treatment effects of behavioral change interventions (Saccardo et al., 2023; Bryan et al., 2021). Brody et al. (2022) show that the efficacy of different communications to increase flu vaccine uptake varies based on individuals' baseline motivation to get vaccinated. Similarly, Homonoff et al. (2021) provide evidence from a large-scale field experiment that student loan borrowers are more likely to reduce late payments and improve creditworthiness when given a personalized credit risk score.

Other studies, motivated by similar research questions, employ different methodological approaches. Some propose the use of selective trials, where individuals' types are identified prior to treatment matching (Dizon-Ross and Zucker, 2023; Chassang et al., 2012). This approach is valuable in principal-agent contexts where incentive alignment may be necessary, but that is rarely the case in public policy. Another stream of studies applies machine learning techniques to big data on individuals' observable characteristics (Athey et al., 2023; Agrawal et al., 2022; Kitagawa and Tetenov, 2018). While promising in many contexts, this process may be challenging to replicate in settings with limited data and stricter privacy rules, as often encountered by government agencies. Our study offers novel evidence that low-cost online surveys, commonly used by public and private sector organizations for routine feedback collection, can provide valuable insights into behavioral barriers to improve intervention targeting.

Insights from our experiment also contribute to the literature on the cost-effectiveness of light-touch behavioral change interventions, particularly those aimed at increasing savings. We make two significant contributions in this area. First, we examine savings in the context of college savings accounts, an area previously understudied, while most available evidence focuses on retirement savings (Thaler and Benartzi, 2004; Laibson, 1997). College savings are an intriguing research avenue due to their shorter and clearly defined time horizons, where suboptimal financial decisions are harder to rectify over time. Additionally, these accounts represent a critical measure of financial investment in human capital, with effective nudge interventions potentially having intergenerational impacts.

Second, the richness of our data allows us to provide evidence on the channels of

nudge efficacy, potentially reconciling mixed findings from previous studies. For example, simplifying retirement plans enrollment process has shown to increase participation rates (Beshears et al., 2013), but encouraging higher contributions into 401(k) plans to maximize employers' contributions had no effect (Choi et al., 2011). Similarly, an experiment in Chile found that assigning savers to groups with their peers with a shared saving goal was effective at increasing savings (Kast et al., 2018), while a field experiment in the U.S. found that sharing information about peer behavior decreased contributions into a 401(k), possibly due to discouragement (Beshears et al., 2015). Mixed evidence also emerges on the efficacy of automatic enrollment, or defaults: while these nudges can act as commitment devices (Ashraf et al., 2006) or anchoring devices (Blumenstock et al., 2018), recent evidence suggests that some effects may be short-lived (Choukhmane, 2019). We offer novel insights by showing that the efficacy of these interventions depends on their relevance to the saver, based on the behavioral barriers they experience. This is especially important in small-sample studies where these barriers may not be equally distributed between control and treatment groups, potentially yielding misleading results.

Lastly, we contribute to the literature documenting how heterogeneous behavioral barriers and beliefs influence behavior. These studies show that erroneous beliefs, even in high-stakes decision-making environments, persist over time without corrective information interventions. Vissing-Jorgensen (2003) demonstrates that differences in investors' self-reported beliefs about future market returns strongly correlate with their portfolio's equity share. Similarly, Barber and Odean (2001) finds that financial traders with biased beliefs lose up to 4% of their annual income. Spinnewijn (2015) and Faberman et al. (2022) show that job seekers' (mis)beliefs about their job-finding prospects reduce job search efforts, leading to long-term unemployment. Other studies using matched surveys to administrative data reveal how biased beliefs influence behaviors in various settings, such as charitable giving (Di Tella et al., 2015), human capital investments (Dizon-Ross, 2019), employers' hiring decisions (Caria and Falco, 2022), and health behaviors (Arni et al., 2021). Our study is among the first to employ this methodology in the context of savings behavior, as previous studies primarily consisted of interventions lacking a behavioral diagnostic phase, such as a

belief-elicitation survey.

The paper is structured as follows: in section 2, we provide a simple theoretical model that motivated our field experiment and that can inform cost-benefit analysis for other personalized nudges interventions; in section 3 we provide a brief explanation of the college savings accounts (i.e., 529 plans). Section 4 explains the variables we collected through administrative records and our survey, details the criteria we applied to identify and measure behavioral barriers in our sample of study participants, and outlines the empirical strategy. In Section 5 we show the pre-experimental descriptive findings and the results of the experiment. Sections 6 and 7 conclude the paper, summarizing our findings and offering research and policy recommendations.

2 Theoretical Framework

We develop a simple framework to guide our hypothesis that personalized behavioral change interventions can be cost-effective. In the Discussion section, we apply the trial results to this framework and derive cost-benefit estimates. Our model is similar to other existing theories of Bayesian updating or behavioral inattention (Gabaix, 2019). We discuss below how our formulation links more directly to policy applications and cost-effectiveness analysis of targeted nudge interventions.

Saver's strategy. Consider an individual i who wants to achieve a savings goal with a degree of uncertainty, subject to a budget constraint.[1] Individual i's savings goal may fall between an optimal T_H and suboptimal T_L amount: $\tilde{T} \in \{T_L, T_H\}$. We define their savings strategy on how to achieve such savings goal as $\theta^i = P(T = T_H)$, and thus their expectation of $\tilde{T}$ can be defined as: $E^i[\tilde{T}] = \theta^i T_H + (1 - \theta^i)T_L$.

The individual's savings strategy θ^i may be suboptimal if their beliefs on how to

[1] In the context of college savings explored in this paper, the source of uncertainty may be the future cost of college or the probability of admission to college for their savings account beneficiary. In a different context, such as retirement savings, the source of uncertainty may be one's life expectancy or future unexpected health expenses.

save are biased. These (biased) beliefs are a function of the signals they receive s and their responsiveness to such signals b - that is, how *relevant* the signals are to i, that is: $\theta^i = \sum_k s_k^i b_k^i$, where $s^i = (s_1^i, s_2^i, \ldots, s_n^i)$ is the vector of signals $1, 2, \ldots, n$, and $b^i = (b_1^i, b_2^i, \ldots, b_n^i)$ are i's responsiveness to each signal. Adopting a canonical generalized hyperbolic discount function (Laibson, 1997; Loewenstein and Prelec, 1992), we can formulate i's utility maximization function as follows:

$$u^i(x) = \underbrace{w - x}_{\text{consumption at } t_1} \overbrace{-\delta_i(\tilde{T} - x)^2}^{\text{distance from the saving goal}}$$

where δ_i is the individual's discount factor of reaching savings goal $\tilde{T}$ *w.r.t.* today's consumption, w is their endowment, and x is their savings rate from forgoing consumption at time t_1. Individual i, who is more or less responsive (b^i) to signals (s^i), chooses a savings rate x that maximizes their expected utility given their information set:

$$\max_x \mathbb{E}[w - x - \delta^i(\tilde{T} - x)^2 | \theta] \text{ where } \theta = s^i b^i$$

We can then derive first order conditions as follows:

$$\begin{aligned}
&\max_x \quad \mathbb{E}[w - x - \delta_i(\tilde{T} - x)^2 | s^i, b^i] \\
\iff &\max_x \quad w - x - \delta_i \mathbb{E}[\tilde{T}^2 - 2x\tilde{T} + x^2 | s^i, b^i] \\
\iff &\max_x \quad w - x - \delta^i \left(T_H^2 \theta^i + T_L^2(1 - \theta^i) - 2x \left(\overbrace{T_H \theta_i + T_L(1 - \theta^i)}^{=\bar{T}} \right) + x^2 \right) \\
\textit{F.O.C.} \quad &-1 + 2\delta(\bar{T} - x) = 0 \\
&\implies x^* = \bar{T} - \frac{1}{2\delta^i} \leq T_H - \frac{1}{2\delta_i}
\end{aligned}$$

The individual i's saving strategy will thus be $\theta^i = \sum_{j=1}^n b_j^i s_j^i$, where $b^i \in [0, 1]^n$ and $s^i \in \{0, 1\}^n$. Based on this model, it follows that signals that better match an individual's beliefs - that is, signals that are more *relevant* - are more likely to lead to a behavioral response.

As a simple illustrative example, assume that individual i receives three signals b_1, b_2, and b_3; their savings decision will be as follows:

$$
\begin{aligned}
& b^i = [0.2, 0.1, 0.7] \\
s^{i'} = [1,1,0]: \quad & \theta^{i'} = b^i s_i^{i'} = 1 \times 0.2 + 1 \times 0.1 + 0 \times 0.7 = 0.3 \\
s^{i''} = [0,0,1]: \quad & \theta^{i''} = b^i s_i^{i''} = 0 \times 0.2 + 0 \times 0.1 + 1 \times 0.7 = 0.7 \\
& \implies x^{'} = (T_H \times 0.3 + T_L \times 0.7) - \frac{1}{2\delta^i} < x^{''}
\end{aligned}
$$

The natural extension of this model is that an individual might receive multiple signals throughout their lifetime, or even actively seek information during important milestones that might lead them to update their savings behavior. However, receiving too many signals may induce greater complexity in the individual's decision-making processes, making it costlier to identify the most relevant one:

$$\theta = sb - sc$$

Social planner's strategy. A benevolent social planner wants to help individuals achieve their behavioral goals by debiasing their decisions using a signalling mechanism. This is often the case in a wide range of policy contexts where a social planner has a limited set of policy levers at their disposal but has the ability to communicate in a low-cost fashion (e.g., email) with individuals who joined their program (e.g., a saving product). In our savings context, the social planner wants to maximize the benefits (i.e., higher savings) of the signals they send to individual i to help them achieve their savings goal. In doing so, the social planner wants to minimize (i) the risks of excessive irrelevant signals and maximize the cost-effectiveness of sending personalized (costlier) signals.

For simplicity, we assume that the social planner knows what the optimal saving strategy is (i.e. knows the realized value of $\tilde{T} = T_H$ without uncertainty) and sends signals s^i

at a cost $c^i s^i = \sum_{j=1}^{n} c_j^i s_j^i$, to individual i:

$$\begin{aligned} \max_{s} \quad & w - x^* - \delta^i (T_H - x^*)^2 - cs \\ \text{s.t.} \quad & x^* = sbT_H + (1 - sb)T_L - \frac{1}{2\delta^i} \\ & s_k(1 - s_k) = 0 \quad \forall k \end{aligned}$$

It derives that in the absence of signaling costs (i.e. $c = (0, 0, ..., 0)$), the social planner's best strategy is to send all possible signals, assuming the recipient i will be able to identify the most relevant and respond accordingly. However, assuming a more realistic context where sending a signal has a non-zero cost, the social planner's utility function can be modelled as:

$$\begin{aligned} (FOC \ w.r.t. \ s_k): \quad & b_k(1 - sb)2\delta^i (T_H + T_L)^2 > 0 \\ & if \ b_K > 0 \implies s_k = 1; then \ s_k = 1 \ \forall k \ s.t. \ b_k > 0 \end{aligned}$$

Thus, the social planner's utility from sending one signal (k) instead of none will be:

$$\begin{aligned} & c_k < \delta b_k [(T_H - T_L)^2 (1 - b_k) + T_L (T_L + 2T_H)] \\ & \quad (call \ d = T_H - T_L \ the \ difference \ between \ the \ optimal \ saving \ rate) \\ \implies & c_k < \delta b_k [(d - 2T_L)^2 (1 - b_k) + T_L (3T_L + 2d)] \end{aligned}$$

From this model it derives that the social planner's decision to invest resources for better diagnostics in order to implement targeted intervention is dependent on the characteristics of the population it aims to help via a behavioral change intervention, in particular their discount factor δ and their observable suboptimal behavior d (e.g., their gap to the optimal savings rate). In the case of savings accounts, as the pool of account owners expands, it is reasonable to assume that the heterogeneity of these characteristics will increase, thus justifying higher costs of targeted interventions, as per our model predictions: $c_k < \delta b_k [(d - 2T_L)^2 (1 - b_k) + T_L (3T_L + 2d)$.

3 Institutional Background: 529 Plans

Higher education is widely recognized as a critical tool for creating equity and achieving economic security. Due to the rapidly increasing cost of college and cuts to public funding for state universities, this goal is increasingly out of reach for many families who are not at the top of the income distribution. College Savings Accounts, also known as 529 plans, are a useful tool for bridging this gap. Named after Section 529 of the IRS Code, these plans are state-administered college savings accounts that allow a contributor either to prepay a beneficiary's qualified higher education expenses at an eligible educational institution or to contribute to an account intended to pay postsecondary educational expenses. In 2001, the Economic Growth and Tax Relief Reconciliation Act made earnings on these plans completely tax-free (with this provision becoming permanent in 2006). Today, all 49 states and the District of Columbia offer 529 plans, and many of them offer income tax deductions for contributions. A recent estimate from 2021 indicates that the total asset value in 529 plans across the U.S. was $480 billion[2].

In Illinois, the state Treasurer (this study's implementation partner) serves as the trustee of two 529 college-saving plans: Bright Start, a direct plan that allows individuals to sign up and manage their savings independently, and Bright Directions, a plan accessible through a financial advisor. Both plans share several defining features: (i) contributions are tax deductible for Illinois taxpayers (up to $10,000 for individuals and $20,000 for joint filers per year); (ii) savings accrue interest tax-free; (iii) withdrawals are tax-free for qualified education-related expenses ; and (iv) assets in 529 plans have a minimal effect on financial aid, with only 5.64% of their total amount being considered as a parental asset. These plans also offer additional advantages: anyone can contribute to a 529 plan and account owners can change the beneficiary or transfer funds to other 529 plans if needed. In Illinois, savers can open an account with no minimum contribution, pay no annual account fee, and are under no requirement to carry a minimum balance. Total fees in the program are also very low. For example, savers selecting the Bright Start Index Age-Based Portfolio (the default choice for

[2]Source: https://www.collegesavings.org/wp-content/uploads/2023/03/CSPN-EOY-22-Data.pdf

Bright Start savers) pay a total fee of approximately 12 cents on every $100 they contribute.

Despite the potential impact and broad availability of these plans, little is known about their take-up and usage, in part because researchers had limited or no access to this administrative data before. Given the novelty of this data, we provide a brief background of the available variables, which will also enter our empirical estimation, in the next section.

4 Data and Methodology

4.1 Administrative data

As part of this study, we have unrestricted access to de-identified administrative data of 529 plans in Illinois both at the account owner level and at the transaction level. From these datasets, we can see that 529 plans have become more popular over time, in large part thanks to the Illinois State Treasurer's Office's allocation of resources to increase awareness among Illinoisan families about the program through, at different times, television and online advertising, community outreach, free online webinars, and direct mail to new parents. Participation in the plans in Illinois went from fewer than 10,000 accounts opened in the program's first year (2000) to more than 46,000 new accounts opened in 2022. In total, about 682,000 accounts have been open in Illinois between 2000 and 2022, which, matched against Census data, equates to about 15% of Illinois children having a 529 plan under their names in the most recent year of data availability. This take up rate is in line with wider trends in ownership of other financial products, such as retirement savings accounts (Census, 2022).

Since the administrative data does not include systematic collection of account owners' income or wealth, we rely on the owners' ZIP code of residence and their total balance across owned accounts to measure the inequality between and within the plans. Descriptive analysis shows that the take-up rate of 529 plans is substantially higher in ZIP codes where the median household income was at least $100,000 than in ZIP codes with lower median incomes; put differently, more than 10% of residents in ZIP codes with a median household

income above $200,000 had an account, compared to less than 2% of residents of ZIP codes with a median household income of $30,000 or less.

While the population of 529 owners appears homogeneous based on ZIP code level characteristics, we note high heterogeneity in savings balances, indicating the presence of inequality not just in take-up rates (between-inequality) but also in savings rates (within-inequality).[3] In 2020, just less than 25% of all money deposited into 529 plans was owned by the bottom 30% of owners, while the top 10% owned more than 50% of it (for a visual summary of these descriptive statistics, see the Online Appendix). This raises the question: are saving disparities entirely driven by income and wealth differences, or do behavioral barriers play a role in hindering higher savings?

4.2 Survey design

In September 2022, we partnered with the Illinois State Treasurer's Office (ILSTO) to launch an online survey for owners of currently open 529 plans in the state of Illinois. ILSTO was responsible, through its outsourced data collection firm Civis Analytics, for sending out the recruitment email to the selected population of Illinois account owners. In order to reach the intended audience, we required that invited account owners had account beneficiaries aged less than 17 listed across all of their accounts and had not opted out from receiving regular electronic communication from the agency, and excluded owners of already closed accounts (see Section 12.3 for a copy of the invite email and the survey). Respondents were compensated with a $5 gift card upon survey completion as well as being automatically entered into a random draw to win an additional $150 for which they could choose how to allocate between another gift card and as a deposit into one of their 529 accounts. We chose a target sample size based on the minimum detectable effect size that the power analysis suggested we should find (see Figure 11 in Appendix A) and the partner agency's budget

[3]As of the end of 2020, the average account balance among open accounts was $26,793 while the median balance was $10,001, and the average account had been open for 6.7 years.

constraints.[4] To reduce the risks of over-recruiting and going over budget, we split the pool of invitees into five subsets to receive the invite email in batches so we could monitor participation rates more closely. The survey remained open for a total of four weeks, reaching a final sample of 3,857 complete responses[5].

Table 1 shows summary statistics of owners' characteristics that are observable from the administrative data, comparing survey respondents to the whole sample of owners of opened accounts. While the differences between survey participants and all other owners who did not participate are qualitatively small (parents with older children and with the highest balances above $500,000 are less represented in our survey sample), we cannot rule out that parents in the former group completed the survey because of unobservable characteristics, such as a higher eagerness to learn something useful about their 529 accounts. A randomized intervention ensures internal validity to control for observable characteristics that may influence savings, such as income (Table 5 in Appendix A shows balance checks between the control and treatment groups).

The median completion time for the survey was six minutes, and 80% of respondents completed it within 14 minutes. The survey was hosted on the Qualtrics platform of the partner agency so that personal identifiable information could be collected and subsequently matched against respondents' records from the agency's administrative database. The matched, de-identified data was then shared with us.

In the subsection below, we explain the rationale behind the survey questions designed to identify the selected behavioral barriers[6], present summary statistics of the survey-administrative matched data, and delineate the criteria for flagging respondents who experienced the barriers. The process of choosing flagging rules for the barriers inherently involves a series of judgment calls. Defining what constitutes a behavioral barrier necessitates

[4]The nature of the survey required that compensation could not come from the researchers, and the University could not reimburse the partner agency for the study costs due conflict of interest.

[5]A total of 112,688 invites were sent, but we have no way of knowing how many emails bounced back or were not opened at all, making it difficult to draw the exact percentage of invitees that completed the survey.

[6]Throughout the paper we refer to "behavioral barriers" as a term to group all sources of seemingly suboptimal behavior that may hinder savings, including behavioral biases, erroneous beliefs, and detrimental habits.

careful consideration. This challenge is a fundamental aspect of any light-touch behavioral change intervention. Previous studies typically overlooked the assessment of these barriers prior to intervention implementation. Our study aims to overcome this limitation by harnessing rich data from both survey responses and administrative records. The unique nature of our dataset allows for the application of nuanced criteria, refining our judgment calls in a series of robustness checks, and evaluating intervention efficacy contingent on previously identified barriers.

4.3 Identifying behavioral barriers

The set of behavioral barriers that influence people's savings decisions are likely to be multifaceted and complex. Using a short online survey as a screening tool requires prioritizing the ones with the most potential to be impactful while ensuring reasonable survey completion times. Additionally, our collaboration with a government agency imposes the requirement to formulate survey questions in a manner that maintains respondent comfort and avoids perceptions of patronization or the implication of a lack of understanding regarding basic economic concepts related to saving. We focused on barriers identified as significant predictors of savings behavior in previous studies. Each subsection below details the decisions behind our flagging criteria.

Expected future cost of college. One critical factor affecting parents' savings decisions for college is their expectation of what college will cost in the future.[7] However, predicting the cost of college in the future is challenging. College websites often display a "sticker price," but the actual cost, considering financial aid, is harder to determine (Levine, 2022). This difficulty is compounded by evidence that many individuals struggle with accurate forecasting, regardless of their educational background or expertise (DellaVigna and Pope,

[7]According to a February 2022 survey by The Princeton Review, 80% of 14,100 respondents, including college applicants and their parents, indicated that financial aid would be "very" or "extremely" necessary to afford college costs (Review, 2022). Similarly, a survey by Discover Student Loans found that 63% of U.S. parents were concerned about having sufficient funds for college in 2021, with this figure rising to 66% in 2022 and 70% in 2023 (Loans, 2023). These findings underscore that college costs are a primary concern for most parents, making their expected future cost a key determinant of how much they save in dedicated accounts.

2018; Tetlock et al., 2017), and it is possible that many parents fail to account for college cost inflation over time.

To measure cost of college expectations, we first asked respondents which type of postsecondary institutions they expected their account beneficiary to attend. Respondents could select multiple options from the following list: 4-year private university, 4-year public university, community college, apprenticeship program, and vocational school. We then posed the following question: *How much do you expect the total cost of undergraduate education to be for the beneficiary of your 529 account—that is, for the whole duration of their degree and including all costs (e.g., tuition, room and board, etc.)?* We deliberately framed the question this way to encourage parents to consider the highest possible cost, including all foreseeable expenses. We compare their estimates against a simple calculation that considers a 5% and 2% annual inflation — two lower bounds based on the trend of the past two decades — and apply it to the cheapest selected institution type selected and account for their beneficiary's years left until college. This allows us to generate a measure of conservative expectation gap, comparing respondents' highest possible subjective expected costs to the lowest possible "objective" future costs[8].

We find that about 41% (29%) of parents fail to account for a 5% (2%) inflation increase when forecasting the future cost of the cheapest institution they hope their child will attend. Among this subset of underestimating parents, the average gap between objective and individual forecast is $79,304 (s.d. $74,589) ($100,690, s.d. $77,988 if accounting for 2% inflation). These estimates increase substantially when considering the most expensive higher education institution type parents hope their child will attend. Figure 4 shows the distribution of this gap. In Table 2, we can see that lower income parents and those with a younger child (i.e., they are further from reaching college age) are significantly more likely to underestimate the future cost of college.

Estimates of future college costs are often shared by education organizations, media

[8]We use sticker price as the objective cost measure because it is the most reliable and universal forecast; while 529 beneficiaries may receive personalized financial aid offers from colleges, such offers are unpredictable, so this approach accounts for all potential scenarios applicable to the entire sample.

commentators, and government agencies. However, it is unclear how savers respond to this information. Would providing parents with more accurate forecasts increase their 529 plan contributions, or would it discourage greater contributions? The lack of empirical evidence on the behavioral response to forecasting corrections makes this question difficult to answer and worth investigating.

Inertia. Defaults are widely used to increase savings, with the assumption that automatic enrollment in recurring contributions helps counteract present bias. Previous studies have shown that automatic enrollment in pension funds leads to higher balances (Thaler and Benartzi, 2004; Ashraf et al., 2006), and defaults can serve as an anchor or signal for appropriate saving amounts (Blumenstock et al., 2018; Madrian and Shea, 2001). However, recent studies challenge this assumption. Choukhmane (2019) finds that employees save less in the future to offset the short-term benefits of auto-enrollment, and Beshears et al. (2024) show that automatic enrollment can lead to increased unsecured debt.

For 529 plans, account owners may opt into a default contribution amount and frequency, which they can modify or cancel at any time. In our context, administrative data show that about 36% of account owners every year utilize an automatic contribution. While this initially increases balances, about half of default users do not adjust their contributions, resulting in lower balances compared to non-default users (see Figures 6 and 7 for descriptive evidence).[9] Although estimating the effects of defaults on 529 plan balances is beyond the scope of this paper, the data suggest that some savers may not make the most of this feature in the long run, possibly forgetting to adjust their default settings, leading to lower balances over time.

Since we cannot predict who will use defaults dynamically, we flag all owners with an active default at the time of the survey as savers at risk of inertia. Table 2 shows that having multiple beneficiaries and marginally higher income are the strongest correlates of default activation. We also find that those using financial planners were less likely to opt into

[9]In a previous 2021 survey conducted by ILSTO, 79% of default users said defaults helped them avoid forgetting and contribute regularly, 50% believed it helped them save more over time, and 35% cited both reasons.

a default, likely delegating their savings decisions and not perceiving a risk of forgetfulness.

Underestimate the power of small increases in contributions. An individual's understanding of compound interest can significantly influence their savings behavior. Lusardi and Mitchell (2011) show that individuals who grasp the effects of compounding interest are more likely to plan for retirement. This understanding is equally crucial for 529 plans for several reasons. First, administrative data reveals that savers typically make an initial large deposit, averaging around $6,000, followed by smaller contributions that add up to about $2,000 a year, which establishes a solid foundation for accruing interest over time. Second, accounts remain open for an average (median) of 7 (6) years, providing sufficient time for returns accumulation. Third, historical data indicates that 529 plans have high annual returns: as of February 2024, accounts with beneficiaries aged 3 to 10 using an age-based aggressive investment option had an average annual return between 8% and 10%.[10] Yet, some savers may underestimate the impact of compound interest on even marginal increases to their contribution amount anyways.

To investigate the presence of this barrier in our sample, we asked respondents two questions: (i) *Do you think you could increase your monthly contributions to your 529 account?* [options: *I could contribute more but don't plan to; I can't contribute more; I can and plan to increase my contributions in the future by $10 to $49, $50 to $99, $100 to $199, more than $200*]; (ii) *Do you think an additional $50 per month to your 529 account would:* [options: *Not make a significant difference by the time my beneficiary goes to college; Make a significant difference by the time my beneficiary goes to college*]. We chose to ask about perceived differences of an additional small contribution instead of testing their knowledge of the compound interest formula since we are more interested in assessing their perceptions rather than their calculus skills.[11]

[10]For a monthly update of the rate of return of 529 plans in Illinois, see: `https://brightstart.com/portfolio-prices-performance/monthly/?portfolio=age-based`

[11]An alternative approach to elicit this barrier would have been to replicate a previously used question in a 2021 survey run by ILSTO where owners were asked: *Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money grow?:* [options: *More than $102, Exactly $102, Less than $102, I don't know*]. Interestingly, 11% of respondents skipped this question, 2% said "I don't know," and 2.5% answered incorrectly. This suggests that a sizeable share of 16% of savers could not answer correctly.

The first notable result is that about 35% of respondents believed that an extra $50 per month would not make a difference by the time their beneficiary began college. To calculate the potential effect of an additional $50 monthly contribution for each respondent, we first estimate the projected savings balance assuming parents would continue saving the same average annual amount they had already been saving and accounting for a compound 5% interest rate (a low-bound estimate of the average annual rate of return in 529 plans). We then replicate the calculation, adding a $50 monthly contribution. Based on these calculations, those who stated that an additional $50 a month would not make a difference would accrue an additional $4,080 on average (s.d. $3,924; min.$600 and max. $14,194). In our main analysis, we flag a respondent as underestimating the power of small contributions if they stated that an additional $50 per month would not make a meaningful difference and our above calculations showed that it would generate an increase in their balance of at least 10% compared to the status quo projection. Doing so flagged 10% of the total sample. In the robustness checks, we relax this criterion and show that the results hold. As per the other beliefs, we regress the dummy variable indicating whether or not a respondent underestimates the power of small contributions on all the available individual characteristics from the administrative and survey data. We find that higher income parents are less likely to hold this misperception, suggesting that higher earners are more appreciative of the importance of small contributions.

Norms perceptions. People tend to follow social norms, but their perceptions of these norms can be skewed without accurate information. They may act according to what they *think* others do rather than the actual behavior of others. Norm misperceptions can hinder behavior change, but interventions providing factual information about norms have successfully corrected beliefs and behaviors in most contexts (Bursztyn and Yang, 2022; Bursztyn et al., 2020). In financial decision-making, however, the impact of norm information is more mixed: experimental studies have shown that providing individuals with information about the prevailing norm can be effective at increasing tax compliance (Hallsworth et al., 2017) but not retirement savings, where it in fact appears to backfire (Beshears et al., 2015).

In our survey, we asked respondents to report their beneficiary's 529 account balance

while providing a hyperlink to their online account login page for reference to ensure both accuracy and awareness of their current savings.[12] Although we could not track respondents' use of the link or their account login times, an embedded timer showed that respondents took about 40 seconds on average to answer, with 80% taking more than 10 seconds, indicating careful consideration of their balance, and possible use of the provided link, prior to answering the question.

Immediately after inquiring about their 529 account balance, respondents were asked to estimate their savings relative to other 529 account holders with beneficiaries of the same age: *Compared to other 529 account owners with beneficiaries of the same age as yours, would you estimate that your current amount of 529 savings is:* [options: *Much higher than average, Higher than average, Somewhat average, Lower than average, Much lower than average*].[13] By matching survey responses to administrative data and calculating beneficiary age-specific savings distributions, we find that, in line with other studies, a sizeable share of 13% of parents erroneously believed they saved more or much more than average. Among these parents, the gap between their savings and the norm was on average $13,929 (s.d. $9,582), with a median of $12,356 (see Figure 5 for a distribution of this gap among this subset of parents). In Table 2 we can see that, unlike the underestimation of future college costs, higher-income parents were more likely to overestimate their saving balances compared to their peers. Additionally, we observe that this misperception diminishes as the beneficiary approaches college age, suggesting that some parents may learn about the true distribution of savings as college enrollment becomes more imminent. It is also worth noting that, compared to other studies that focus on social norms, in our specific context there is not a normative norm that parents should follow, but rather a social norm as a reference point parents might find useful to adjust their saving goal.

Given the existence of norm misperceptions and the inconclusive evidence regarding the effectiveness of norm-correcting information treatments in this context, addressing this

[12]The question asked: *"What is the approximate balance of your beneficiary's 529 account? If you'd like, you can check the amount by logging in here."*

[13]An alternative approach would have been to ask respondents to provide a figure of what they believed was the average balance of other comparable parents. However, we were more interested here in the perception of the norm, rather than the accuracy of the guess.

barrier presents an ideal opportunity to estimate the treatment effect of targeted behaviorally-informed communication.

In this section, we defined the methodology for identifying and flagging whether an individual experienced one of the selected behavioral barriers. In the next section, we explain the experimental component of the survey.

4.4 Randomized intervention

The final page of our survey contained a randomly assigned behaviorally-informed message that encouraged account owners to increase contributions to their 529 accounts. An important but often overlooked challenge with behavioral change interventions is whether it is welfare maximizing – that is, whether improving a person's outcome in one dimension comes at the expense of worsening their outcome elsewhere. In the scenario that our treatments were effective at increasing contributions into 529 accounts, one would want to know if this came at the cost of lower contributions into other savings accounts, such as retirement plans, or affected a person's financial well-being.

In the case of 529 accounts, higher saving balances typically do not negatively affect FAFSA eligibility because they are considered parental assets and assessed at a relatively low rate of 6% in the financial aid calculation. Nevertheless and as we do not have access to household savings and consumption beyond what we can track in their 529 accounts, we minimize the risk of negative spillovers by asking respondents two direct and simple questions early in the survey: *Do you think you could increase your monthly contributions to your 529 account?* [options: *I could contribute more but don't plan to; I can't contribute more; I can and plan to increase my contributions in the future by $10 to $49, $50 to $99, $100 to $199, more than $200*]; and *If you were to increase your monthly contributions into your 529 account, what trade-offs would you have to face?* [options: *I wouldn't face any substantial trade-offs; I would reduce my discretionary spending; I would reduce my spending on necessities; I would save less elsewhere (e.g., retirement, stocks, etc.); I would accrue debt*

elsewhere]. In addition to minimizing the risks of negative effects, this question also allows us to be closer to our goal of identifying a sample of savers for whom the treatments might be relevant.

We embedded display logic into our survey instrument that automatically skipped displaying a randomized final treatment message for respondents who answered that they could not contribute more in the first question or that they would have to reduce spending on necessities, save less elsewhere, or accrue debt elsewhere. These respondents, which account to 26% of the sample (n=985) were not part of the randomization and saw instead a final message identical to the control group. Further, we implemented display logic to exclude individuals flagged as 'unbiased' (around 25% of the total sample) based on predefined criteria established in collaboration with the partner agency. This decision aimed to mitigate potential backfiring effects that may lead to complaints or account owners' dissatisfaction. This subset of 'unbiased' owners saw a social norms message (see description below). Both the financially constrained and unbiased subsets were not included in the randomization, and thus, we exclude them from the main outcome analysis. Image 10 provides a visual summary of the exclusion rules and the random assignment.

Section 10.1 of the Appendix details the control and treatment messages used in the study. The control group message did not include any encouragement or behaviorally-informed content; instead, it reminded respondents of the free online self-help resources offered by the government agency, which they could access at any time. These resources include a free online calculator to estimate future college costs and a description of the impact of small contributions due to compound interest. We redirected control respondents to this webpage to ensure they were not disadvantaged compared to the treated respondents and had access to most of the same information, albeit without the behavioral framing. The "adjusting defaults" treatment message explained the automatic transfer feature, ensuring respondents were aware of its availability and reminding them of the importance of regularly reviewing the frequency and amount if they were using it. The "saving norms" message provided respondents with a table showing the distribution of savings balances across quintiles and by beneficiary age, adjusted to the age group of the respondent's beneficiary (1-5, 6-9, 10-13, and

14-17). To address the underestimation of future college costs, we presented randomly selected respondents with an age-adjusted, college-type specific table showing the forecasted cost of college attendance, accounting for 5% inflation (a conservative benchmark based on trends of the past two decades). Lastly, the "power of small increases in contributions" message included a table illustrating the effect of small monthly contributions over a four-year time period with a 5% compound interest rate (also a conservative estimate based on past trends for 529 plans). Previous studies have demonstrated that a simple calculator can effectively explain the power of compound interest on savings (Hubbard et al., 2016), so the table was designed to be a user-friendly tool for account owners.

To ensure the treatment messages were as comparable as possible, all messages included the following wording: *A small increase in the amount of your 529 recurring savings today can make a big difference tomorrow. Don't delay such an important decision. Review now the amount and frequency of your 529 contributions to make sure you reach your saving goals. You can log in to your account by clicking here.* The purpose of this message was to minimize any potential discouragement and provide a clear call to action. Additionally, to eliminate any administrative burden and make it as easy as possible for respondents to contribute to their accounts, we embedded a hyperlink on the word "here" that redirected respondents directly to their account login page.[14]

4.5 Empirical Strategy

In this subsection, we explain how we evaluate the effects of our randomized treatment messages. The main outcome of interest in our analysis is the difference in savings balance before, at time *t*, and after, at time *t+1*, completing the survey, in line with previous studies (Ashraf et al., 2006). The choice of the main specification is informed by three important considerations related to our data and research question. First, our outcome variable doesn't take on negative values: owners with beneficiaries not yet enrolled in college rarely withdraw

[14]If respondents clicked on the link a new browser window would open and they would land on their account login webpage: `https://portal.ubtrust.com/Portal/paport/brightstart/#login`

money before college enrolment since withdrawals are only allowed for education-related expenses (and would be heavily taxed otherwise). We don't see any withdrawal from our sample over the observation period. Second, to control for possible negative spillover effects where any increase in savings in one account might come at the expense of lower savings into another account, such as that of a second child, we focus on the average balance, thus dividing the sum of all balances across owned accounts over the number of owned accounts. The number of owned accounts still enters our specification among the other covariates. Third, we are interested in capturing decreasing marginal utility over the outcome: it is reasonable to assume that parents' objective function is to save for their child's college up to a certain point after which they might derive greater utility by saving elsewhere (e.g., retirement). We thus estimate the following model:

$$\Delta Y_i = \alpha + \beta y_{i,t} + \gamma T_i + \delta X_i + u_{i,t} \tag{1}$$

The dependent variable ΔY_i is the owner i's difference in account balance between the pre-trial period *t-1* and one year later.[15] The coefficient β captures the lagged dependent variable, which is the pre-treatment balance y at time t. The importance of controlling for the lagged dependent variable is threefold: (i) it captures the marginally decreasing utility from additional savings; (ii) it accounts for heterogeneous starting balances, acknowledging that the same amount increase over a year has different implications based on the owner's starting balance; and (iii) it minimizes the risk of omitted variable bias since starting balance may be a valuable proxy for wealth and disposable income (Wooldridge, 2010)[16]. In Section 11.0.1 of the Appendix we report the results of a set of tests that allows us to rule out autocorrelation and endogeneity. The coefficient γ is the ITT of each randomly shown message T. We then control for a rich set of covariates X using both survey and administrative data, and

[15]The timing of the data pulls were previously agreed upon with the institutional partner to have enough granularity of the outcome variable while at the same time minimize administrative burden on the government agency.

[16]An alternative specification would be to regress the outcome variable at time t (instead of the difference over time) on the lagged dependent variable. While the magnitude and interpretation of the coefficients changes, the statistical significance of the main effects remains unchanged.

account for the error term u.[17] We estimate this model using generalized linear model using a log link function with gamma distribution and robust standard errors. This model is the most suitable since our outcome variable is continuous, cannot take negative values, and exhibits right-skewness due to the savings inequality previously documented in the data (see Figure 22 in Appendix) (Chen and Roth, 2023). This also allows us to more easily interpret the coefficients in terms of percentage changes with heterogeneous starting balances. The inclusion of the covariates X, while balanced between control and treatment (see 5), is useful to qualitatively compare the results when we match treatment to behavioral barrier in order to estimate the effect of seeing a relevant (or irrelevant) message.

5 Results

5.1 Pre-experimental results

We begin our analysis by examining the correlation between the selected barriers and savings balance prior to exposure to the treatment. Our first notable finding is that the barriers are not systematically correlated with observable characteristics and are not strongly correlated with each other (see tables 4 and 2). About 25% of respondents did not seem to experience any barrier, 44% only one, 25% two, and about 6% displayed three or four barriers. These are notable percentages that point to a previously unobservable heterogeneity among a population of account owners who appear rather homogeneous on observable characteristics.

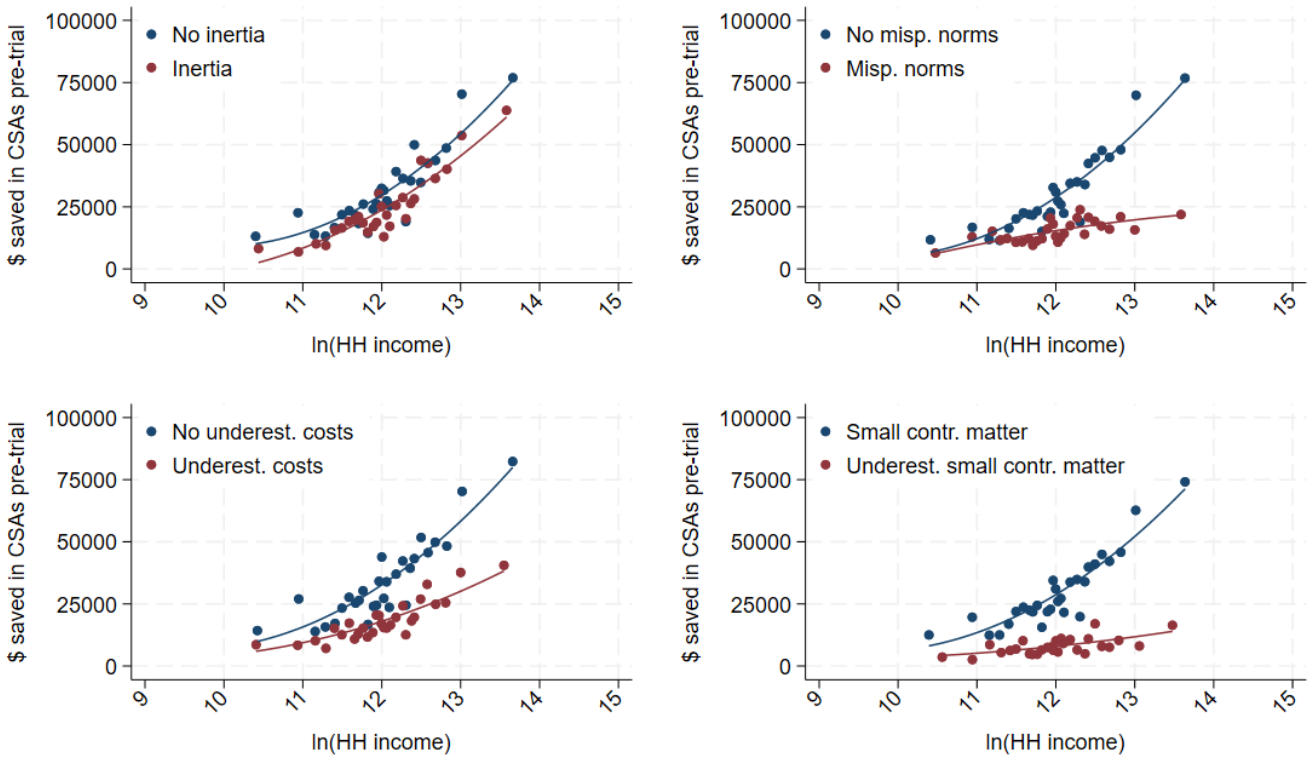
Result 1. Behavioral barriers are widespread but most savers display only one or two barriers and are not systematically correlated with observable characteristics.

We then look at whether these barriers predict lower savings (prior to treatment exposure). From Figure 1, we can see that each barrier is correlated with lower savings

[17]Table 19 in Appendix reports the results of a post-estimation test demonstrating the low level of multicollinearity across independent variables.

regardless of household income level. This suggests that parents' biased beliefs may suppress savings for their child's higher education regardless of their socio-economic status.

Figure 1: Biased beliefs and savings, across incomes

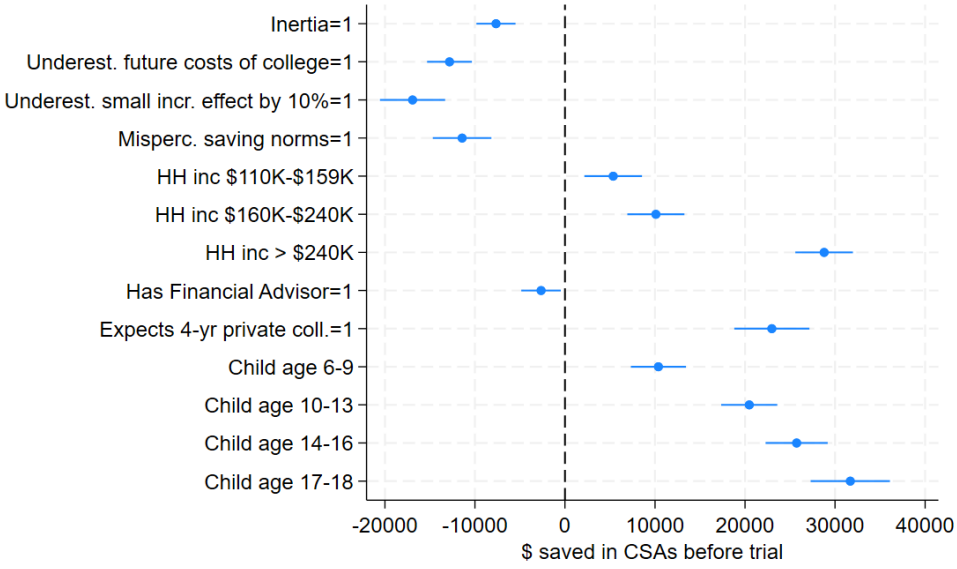


Notes. The four binscatters show how each biased belief is associated with saving balances along the income gradient. Each graph shows the relationship between savers' household (log) income on the x-axis and the balance across their savings account a month prior to the experiment on the y-axis, absorbing for expected college choice.

Figure 9 in Appendix B plots the coefficients of a linear probability model with savings balance as the dependent variable regressed on observable characteristics. This analysis confirms that, while savings are higher among higher income households and among those with older children, experiencing any barrier is correlated with lower saving balances by an average of about \$14,000 (from \$7,661 up to \$16,000).

Result 2. Behavioral barriers are correlated with lower savings after controlling for saving goal and a rich set of characteristics, including income and education.

Figure 2: Correlates of saving balances (pre-trial)



Notes. The image shows the coefficient estimates and the 95% confidence intervals of correlates based an OLS regression with saving balance pre-trial as dependent variable.

We next estimate the average treatment effects of the behaviorally-informed messages compared to the control group, using Equation 1.[18]

5.2 Experimental results

Simple specification. Results from Table 6 show the average treatment effect of each message compared to the randomly assigned control group. Account owners who received the

[18]All regressions follow the GLM model specified in Equation 1 unless another model is explicitly defined. We exclude 157 accounts (∼6% of the sample) that closed within a year after the survey. The administrative database is set up in a way that balances of closed accounts are overwritten as $0, thus not being able to track the latest balance prior to account closure. The closure rates were similar between the control (7%) and any treatment group (6%), t=1.2201 (df=2,389) p=0.222

message encouraging the use and review of automatic contributions increased their saving balances by a range between 16% and 29% compared to the control group, equivalent to $877 to $1,603 more after one year (or $73 to $134 more per month). We observe a similar effect among owners who saw the message that detailed the effect of compound interest, with balances 16% to 19% higher than the control group. Conversely, messages on saving norms and future college costs appear to have backfired: the saving norms message, in particular, underperformed by up to 12% compared to the control group, meaning these owners saved about $55 less per month than control group owners.

Thus far our trial resembles previous studies testing the efficacy of behaviorally-informed communication. However, failing to take into account how relevant these messages were to the receivers might encourage the policymakers to scale up seemingly effective messages and drop seemingly ineffective ones. In the next section, we replicate the regression taking into account the relevance of the message and show that stopping at a simple specification without accounting for this relevance-heterogeneity would not have been the best strategy for the social planner.

Heterogeneous treatment effects: pooled analysis. In this analysis, we pool all the treatment messages into two categories: "relevant", which are the cases where the individuals saw a message that matched a behavioral barrier they experienced, and "irrelevant" as the cases where individuals saw a treatment message without experiencing the barrier that the message aimed to correct. This allows us to see what would happen if a social planner replicated the intervention by targeting each message to its intended recipient.

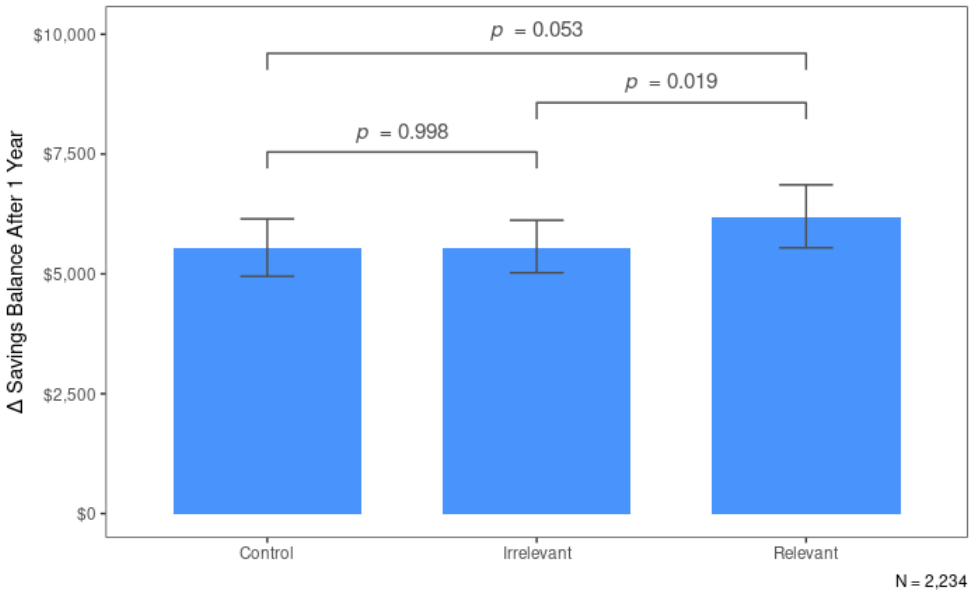
Extensive margin. As part of the data pulls agreed with the partner agency, we were able to track actions taken on savings accounts within 30 days from survey completion. From Table 8, we can see that respondents who saw a relevant message were on average around 16% more likely to make a deposit within this time frame compared to the control group, while those who saw an irrelevant message were about 14% less likely compared to the control group to do so. It is worth noting how the friction costs of logging into the online account to make a deposit were likely low for all respondents. In one survey question

where we elicited respondents' current savings balance we provided them with a link to the login page of their account. About 17.5% of respondents spent more than one minute on this question, suggesting that some might have checked their personal account in another browser tab while completing the survey. Additionally, completing the survey may have put all account owners in the mindset of reviewing their contributions. The fact that we see differences across groups suggests that the message relevance may have been what prompted account owners to take action and make a deposit.[19]

Intensive margin. Looking at savings behavior (see Table 7) we see that a relevant message marginally increased saving balances after one year by an average of 7% to 11%, equivalent to an additional $33 to $55 on average per month. The relevance effect compared to control is marginally significant in the basic specification (p=0.084), but becomes more imprecise with the inclusion of controls (p=0.196) and fixed effects (p=0.336). An irrelevant message performed, on average, worse than the control group, although the difference is statistically insignificant. A post-estimation equivalence test between the relevant and irrelevant variables show a large and significant difference across all specifications ($\chi^2 = 4.65$, p=0.00012 in the full specification with controls and fixed effects), suggesting that the same messages had a significantly different effect depending on whether they were relevant or irrelevant to the recipient.

[19]We explore an alternative explanation using the survey completion time as a proxy for attention. While we didn't measure the time spent on the messages themselves, the total amount of time spent completing the survey can be an indicator of general interest in learning more about college savings. Regression results reported in Table 17 show no significant difference between relevant and irrelevant messages on completion times, and no significant difference between these two treatments against control, partly ruling out differences in attention or engagement as an alternative mechanism.

Figure 3: Message efficacy by relevance



Notes. The graph shows the average difference, and the 95% confidence interval, in savings balances after one year between the control group and the interventions pooled between relevant and irrelevant.

These results support our main hypothesis that the efficacy of light-touch behavioral change interventions is conditional on its relevance to the recipient.

Result 3. Behavioral change intervention efficacy varies by its relevance to the recipient.

The richness of data also allows us to go a level deeper and assess the extent to which the relevance effect differed across behavioral barriers. It is possible that some barriers may be harder to overcome with a light-touch communication intervention.

Heterogeneous treatment effects: barrier-specific analysis. The heterogeneous treatment effects analysis based on individuals' experienced barriers is conditional on our sample size (which is dictated by our partner budget constraints): since each owner might experience multiple barriers, the total number of combinations of barriers by the five possible messages individuals could see (four treatment messages and one control) yields 80 possible interaction effects. Such an analysis would not yield insightful results given our

sample size. To overcome this constraint, we employ two approaches. The first approach consists of running regressions on the subset of owners who saw either the control message or one of the treatment messages, interacting the treatment message with each barrier. This is in lieu of doing a series of pairwise comparison t-tests, accounting for model restrictions and all available controls. The main advantage of this approach is that we can safely rely on the balanced randomization between the control and each treatment message and not require a large sample size to compare each treatment group against each other. The second approach subsets the data by including only those who held a specific barrier (e.g., only those who misperceived social norms), then estimates the treatment effects of each message within this subgroup and using a continuous variable for the barrier (e.g., the gap between one's savings balance and the norm) for the interaction terms. The advantage of this approach is that while controlling for whether the owner experienced other barriers, we do not need to account for all possible combinations of barrier-treatment and can simply identify which treatment message worked best for each barrier experienced.

Future cost of college. We begin our heterogeneity analysis by focusing on a message that appears to have backfired in the basic specification: the projected costs of college. We run a regression on the subset of respondents who saw either the control group message or the future cost message and interact the treatment message with a dummy indicating whether the account owner did underestimate the future costs. The change in signs of the interaction term suggests that the backfiring effect might dissipate when the message is relevant to the recipient, although it remains statistically insignificant (see column (2) in Table 15). Next, we explore whether the relevance of the treatment message is conditional on the extent to which an account owner's subjective estimates differ from government forecast. For consistency with the norms message heterogeneity analysis, we create a categorical variable using the terciles of the gap between subjective and objective costs (the omitted category in column (3) of Table 15 flags respondents who overestimated future costs (59% of the sample)). The interaction term shows that the relevant message reverses this initial backfiring effect for individuals whose estimates were off by up to about 49%, but there is a negative, although not statistically significant, effect for those whose estimates were off by a large margin. Similar to

the norms analysis, it is possible that account owners who learned they were overestimating the future cost of college (the omitted category in our regression) reduce their savings effort. These results confirm our hypothesis that the efficacy of a behavioral change message is conditional on the individual's prior beliefs: account owners saved more when presented with accurate forecast costs, as long as their underestimation wasn't off by a margin that was too large to close (see Figure 15 for a visual summary).

Analyzing the effects of each randomized message on the subset of respondents who underestimated the future cost of college in column (2) of Table 16, we see the same effect: the future cost of college message was most effective among owners who underestimate it by a manageable margin. We also see, however, that owners who underestimate the future cost of college were responsive to other treatment messages too, compared to the control group. This might suggest a demand for information by parents who are least knowledgeable about future costs and have a propensity to change their saving behavior when nudged.

The power of small contributions. Next, we examine the heterogeneous treatment effects of the compound interest calculation message. Column (1) in Table 13 confirms that holding the misbelief that small contributions do not make a significant difference (when they could actually increase balances by at least 10% more than status quo projections) is correlated with lower saving balances. The interaction term between the message and the biased belief dummy is statistically significant and negative. In column (3), we follow the same approach as in the previous analysis and interact the treatment message with a categorical variable that takes tercile values of the gap between status quo future balance and projected balance with an additional $50 per month (see Figure 14). The interaction terms are positive, although not significant. Our interpretation of this result is that underestimating the power of small contributions due to compound interest is a harder biased belief to shift. The heterogeneity analysis comparing the efficacy of a relevant message against irrelevant messages (shown in Table 14) seems to support this claim: one of the treatment messages proved more effective than the control group message at increasing savings. One possible reason is that account owners might perceive each contribution in terms of a share of the final savings goal or as a share of the future cost of college they aim to cover, thus perceiving

small contributions as insignificant, discounting the power of compound interest over time.

Adjusting defaults. Lastly, we analyze the heterogenous treatment effects of the adjusting defaults message. Table 9 shows that this message was effective overall at increasing savings balances by up to 25%. We also see from the administrative data that those who saw this message marginally increased the amount of their automatic contributions (by about $5 more on average, t=1.538, p=0.062). This indicates that the balance increase may be partly due to changes in regular contributions settings as well as higher lump sum contributions. This effect was not particularly more pronounced among those already using defaults, nor did it vary based on how long they had being using this setting (which could have signalled potential forgetfulness). Figure 12 provides a visual summary of the results.

Column (1) in Table 10 reports the results of a regression on the subset of owners flagged as possible procrastinators, confirming that the message encouraging adjustments to default settings was the only one that did not backfire among this group of account owners.

Saving norms. We conclude the heterogeneity analysis by focusing on the efficacy of the saving norms message. Column (1) in Table 11 confirms the main specification results: showing saving norms, on average, backfired, leading to lower savings compared to the control group. However, the interaction term between the treatment message and the dummy indicating whether the owner did in fact hold norm misperception beliefs in column (2) is positive, statistically significant, and has a large coefficient. This suggests that the peer comparison message was effective at increasing savings rates among those who misperceived the norm savings amount, but backfired among those who didn't misperceive it or underestimated how much they saved compared to others. Thus, the backfiring effect detected in the initial regression appears to be driven by those who found the message irrelevant or too reassuring about their savings balance. This effect does not appear to be sensitive to the magnitude of such misperception.[20] Figure 13 provides a visual summary of this result.

[20]To explore an additional level of heterogeneity, we create a categorical variable using the terciles of the (continuous variable) gap between one's savings balance and their peers' average. Column (3) of the same table show that the backfiring effect is not sensitive to this self-others gap.

Using a second approach to estimate heterogeneous treatment effects, we run a regression only on the subset of owners who held norm misperceptions to identify which message was most effective (see Table 12). Other treatment messages, such as adjusting defaults and the power of compound interest, were similarly effective at increasing savings regardless of the gap between one's savings and the average of comparable peers. Our preferred interpretation is that norm misperceptions of savings are an indicator of owners' lack of reference points for a sufficiently high savings balance, and thus almost any encouragement message might be effective. The opposite is also true: information treatments that correct norm misperceptions might backfire if they confirm a person's belief that they are saving more than others.

5.3 Crowding out

One important, but often overlooked, limitation of trials studies is that individuals' decisions are often observed in isolation. It is possible that an increase in contributions into college savings accounts crowded out contributions into other savings accounts, such as retirement accounts. Unfortunately, our trial is not immune from this limitation either since we do not have data sharing partnerships with all possible retirement savings providers. However, we partly overcome this limitation by screening out survey respondents who stated in the first survey module that they would be have to face substantial trade-offs.

Another possible source of crowding out is in other, non-financial, forms of human capital investment. It is possible that additional deposits into 529 plans lead parents to invest less in other activities, such as time spent with their children. On the other hand, a crowding in mechanism is also possible if contributions into college savings accounts act as a commitment device for parents to help their children pursue higher education goals. To investigate this mechanism, we analyze respondents' preferences for allocating the $150 lottery incentive. This preference was elicited on the final page of the survey, after showing respondents the control or treatment messages.[21] More than half of respondents opted to

[21]The lottery question asked all respondents: *Thank you for your participation. This is the last question*

receive the entire amount as a deposit into their accounts instead of a bookstore gift card. As report in Table 18, we found no significant difference across the three groups (control, relevant, and irrelevant) in the amount allocated to the deposit versus the gift card, with both the relevant and irrelevant groups allocating an average of $112 as a deposit (t= -0.1352, df=1901, p=0.892).

6 Discussion

In this study, we utilized a short survey as a belief-elicitation tool to assess the cost-effectiveness and scalability of personalized behavioral change interventions addressing specific beliefs. Our findings indicate that relevant messages have a more substantial impact than irrelevant ones and perform better than redirections to online self-help resources (our control group). The central question remains whether this tool is cost-effective and scalable. To explore this, we applied the scalability framework proposed by List (2022).

Costs. Public and private sector organizations frequently collect feedback through surveys. Our study shows that a concise set of belief-eliciting questions can effectively identify biases and measure their impact on behaviors when aligned with administrative records. Therefore, organizations can potentially leverage existing data collection infrastructures at zero marginal cost by incorporating additional questions to deliver targeted interventions. Furthermore, DellaVigna et al. (2022) argue that nudges are more likely to scale when they involve minor changes to existing architectures. Since many organizations already have the capability to conduct regular surveys, our approach is relatively easy to adopt.

of the survey; once you answered, please click the 'next' arrow button to submit your responses. In addition to receiving the $5 voucher, you have been entered into a lottery for the chance to win a $150 online voucher towards purchases at bookseller Barnes & Noble. If selected as the prize winner, you can spend the voucher on any item(s) of your choice at Barnes & Noble, including books, textbooks, stationery, and other items that might be useful towards your beneficiary's education. You may also choose to split the voucher to receive none, some, or all of the winnings as a direct deposit into your 529 account. Please select below the amount of money you would like to subtract from the $150 Barnes & Noble voucher that would be deposited into your 529 account in the case you are selected as the prize winner. Should you win, the amount requested as a contribution to your 529 account will be deposited directly into your 529 account in about four weeks, and the portion you chose to receive as voucher will be sent via email to you from Civis Analytics in about four weeks.

To mitigate self-selection bias, we rewarded survey participants with a small $5 participation fee and a chance to enter a $150 lottery. Future studies might explore lower-cost compensation schemes, such as using only a higher amount lottery reward. Organizations can also utilize other feedback collection points, like during the signup process for an online account or service, or by employing new AI tools like chatbots through pop-up windows, further lowering implementation costs.

Benefits. Personalized behavioral change interventions may offer greater benefits than traditional random allocation approaches. A screening survey instrument, like the one used in this study, minimizes negative spillover risks by avoiding crowding-out effects. This was evidenced by our investigation into respondents' trade-offs for increasing their 529 plan contributions. Additionally, such instruments enhance the fidelity of the intervention, ensuring consistent quality when implemented at scale and requiring minimal supervision compared to other interventions that might need new data collection infrastructure or skilled human resources.

Using a simple cost-benefit ratio to evaluate our experiment, we assumed incentive costs at $5 per person, plus an estimated staff cost of $100,000 (if the partner agency hired a private firm to conduct the experiment instead of using pro-bono researchers). Considering a mid-point estimate of $62 extra per month per saver treated with personalized messages (n=636), after one year, the intervention yielded a cost-benefit ratio of 1:4. The benefits will likely be substantially higher in the long run due to compound interest. An extra $62 monthly contribution for five years, assuming a conservative 5% compound interest rate, would yield an additional $3,723 on average in saving balances. Thanks to the low-cost scalability of our intervention, the cost-benefit ratio in the medium term justifies even substantially higher individual incentive costs to collect beliefs data on a larger and more diversified pool of savers, as hypothesized in our theoretical model outlined in Section 2.

Limitations. Like any behavioral change intervention, our study has limitations. One major limitation is the difficulty in recruiting a diverse pool of participants for personalized interventions. Although our theoretical framework and cost-benefit analysis suggest that high

incentives to reduce self-selection bias are justified, this remains a significant challenge.

Additionally, a social planner might prioritize those who opt into such surveys, as their participation indicates a demand for personalized advice that should be met with targeted, relevant information.

Another limitation is that we did not incentivize the accurate elicitation of barriers or beliefs. For instance, we could have rewarded survey respondents for accurate forecasts of college costs or compound interest rates, as done in other studies (Haaland et al., 2023; Fuster and Zafar, 2023). While rewarding correct responses could improve accuracy, it might deter participation if parents perceive it as a knowledge test, which our partner agency wanted to avoid.

Lastly, labeling behavioral barriers always involves a degree of judgment. Although we provided comprehensive evidence to support our decisions and took a conservative approach to minimize interpretation bias, this remains an inherent limitation. Behavioral change interventions without pre-collected barriers data often rely on assumptions about behavioral responses based on researchers' qualitative judgment rather than evidence. Our study underscores the importance of embracing transparent exploratory approaches to better understand human behavior before implementing large-scale behavioral change interventions.

Policy Impact. The Illinois State Treasurer's Office, as the implementation partner for this experiment, effectively integrated key learnings from the project to establish a robust infrastructure for delivering personalized behavioral interventions at scale. Among the new service improvements offered to owners of 529 plans in Illinois, the agency built on insights from the experiment to help owners overcome the backfiring effects of inertia by introducing a feature that allows owners to opt in for automatic increases in their contribution amounts over time (see Figure 18). Additionally, the Treasurer's Office launched a new app designed to simplify the process for parents to contribute to and track their 529 savings plans. The app's homepage is specifically tailored to deliver randomized, personalized messages to users, which will be routinely tested for effectiveness in motivating higher engagement (see Figure 19). This system ensures a continuous cycle of experimentation and learning, allowing the agency

to refine its approach and deliver more impactful, data-driven, personalized interventions.

7 Conclusions

This study evaluated the effectiveness of personalized behavioral change interventions on increasing savings in 529 college savings plans. We aimed to identify cost-effective, scalable strategies to improve savings behavior. Through a field experiment, 529 account owners participated in a survey that elicited their beliefs about college costs, compound interest effects, saving norms, and preferences for automatic plans. Participants were randomly assigned to receive one of four behaviorally-informed messages or a control message offering online self-help resources.

Our results show that messages significantly influence savings behavior when addressing the recipient's specific behavioral barrier. Our findings suggest that personalized interventions can enhance savings behavior cost-effectively, especially when implemented through existing feedback mechanisms. Organizations could embed behavioral barrier-eliciting questions into routine surveys or chatbots to address biases at minimal cost, ensuring consistent quality when scaled. A cost-benefit analysis indicates that personalized messages deliver substantial long-term benefits, justifying higher participation incentives and broader implementation. Policymakers should consider integrating personalized communication into financial education programs and savings plan communications to maximize their impact.

A similar approach could be effectively applied in other policy areas such as healthcare and education. In healthcare, personalized nudges could be integrated into patient communication systems to encourage healthier behaviors, such as medication adherence, preventive screenings, or lifestyle changes. For instance, patients could receive tailored reminders based on their specific health conditions, beliefs, or behavioral barriers, increasing the likelihood of positive health outcomes. In education, personalized communication could be used to support student engagement and performance, such as through customized messages that address common obstacles to learning (e.g., inattention, motivation).

Compared to traditional light-touch behavioral change interventions, this approach acknowledges the need to understand the source of seemingly sub-optimal behavior, such as incorrect but malleable beliefs or other behavioral barriers. It also mitigates backfiring effects when scaled, thanks to data that explain why an intervention might not have worked. Ultimately, our study underscores the importance of embedding greater humility in behavioral change interventions to account for individual motives and behavioral barriers, which would improve the impact of these interventions.

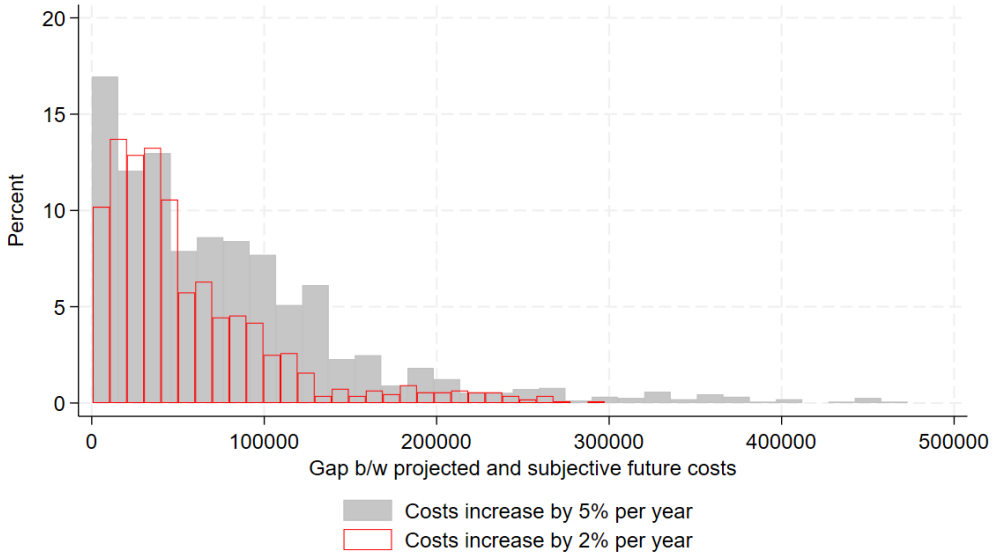
8 Appendix A - Descriptive evidence of behavioral barriers

Table 1: Sample characteristics

	Survey participants	All other owners	Diff.
Savings balance pre-trial			
Mean	28,351	23,070	-5,281
S.d.	39,960	38,321	-1,639
Min	0	0	0
Max	519,586	1,084,575	564,989
N	3,825	294,210	290,385
Other observable characteristics			
Mean (oldest) account lifespan	6.2	7.8	1.6
Share who lives in 4th or 5th income quintile IL zipcode	59.9%	63.7%	3.8%
Mean age oldest account beneficiary	10.0	14.9	4.9

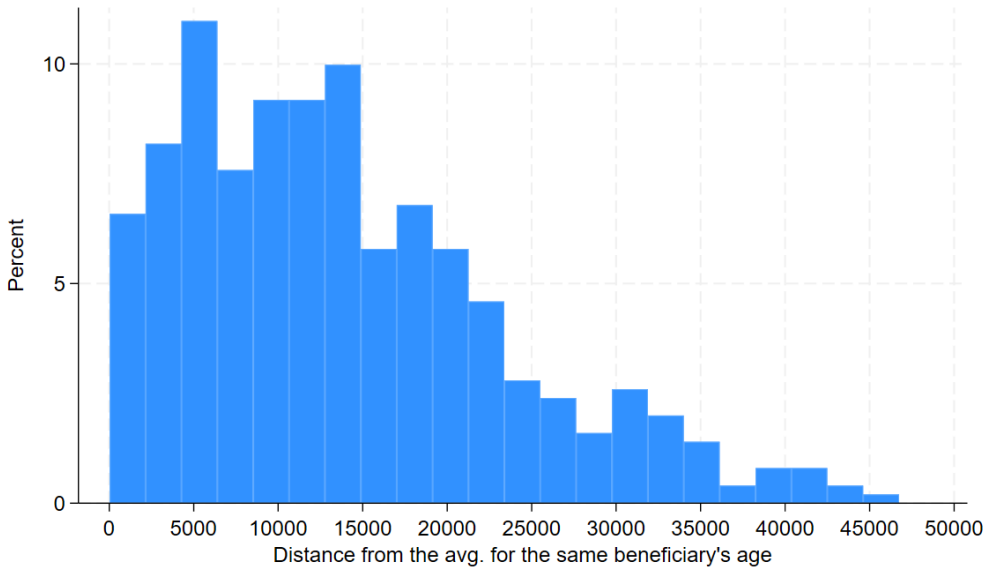
Notes. The table shows the characteristics of the survey participants compared to the rest of all other account owners who did not take part in the survey.

Figure 4: Underestimation of future cost of college



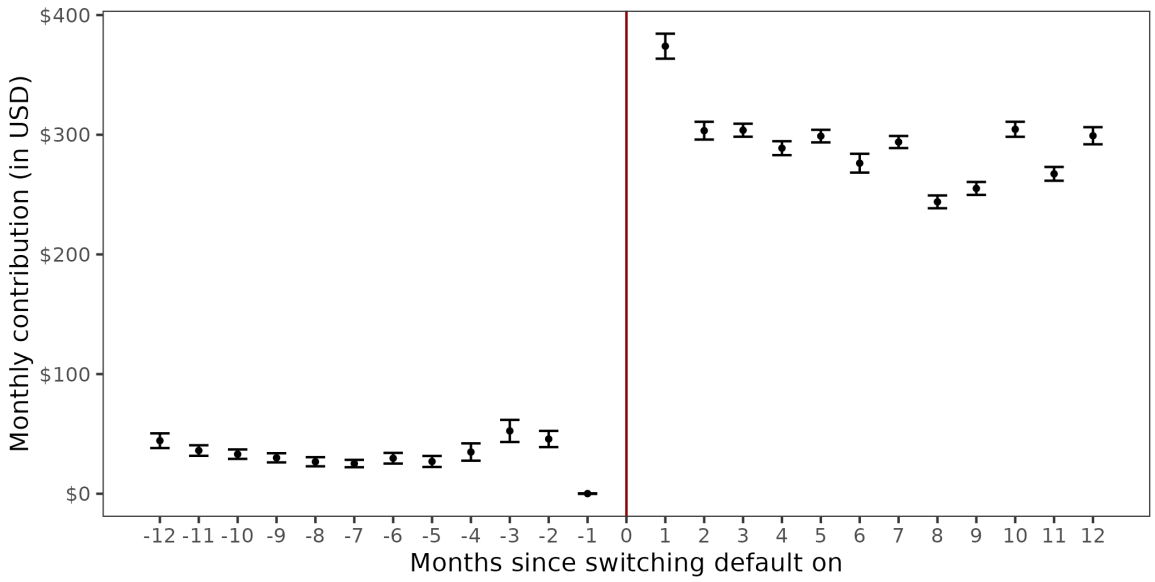
Notes. The image shows the percentage distribution of the gap between a 2% and 5% projected cost increase and respondents' forecast for the cheapest higher education type they expect their beneficiary to attend.

Figure 5: Norm misperception



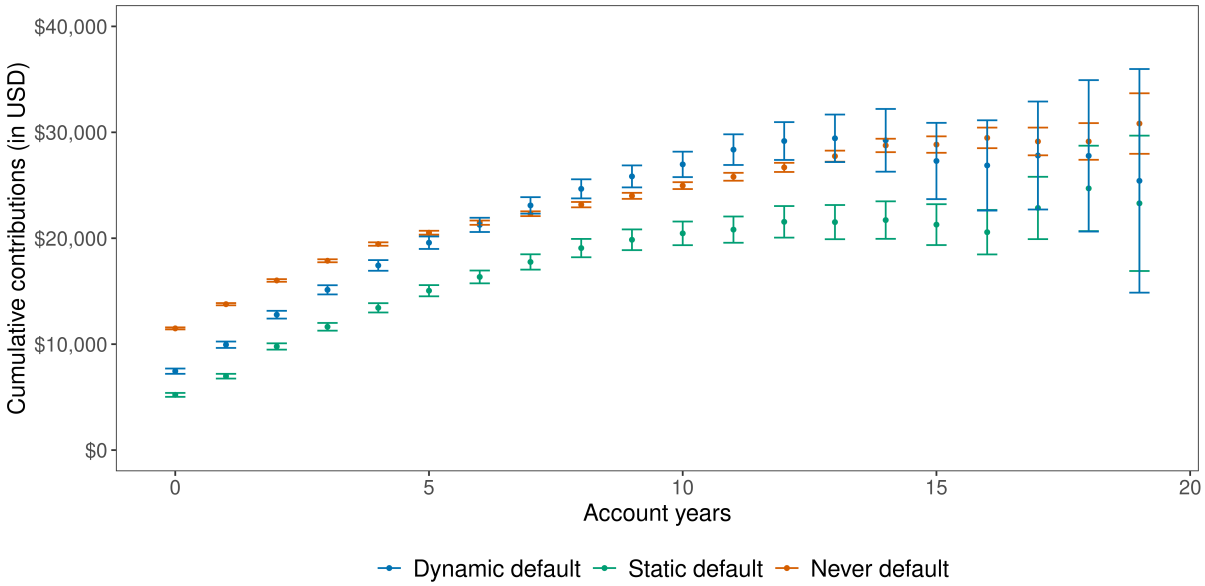
Notes. The graph shows the distribution of the distance between the respondents' savings balance and the average balance for owners with beneficiaries of the same age, using the average of currently open accounts only.

Figure 6: The short-term effects of defaults on saving contributions



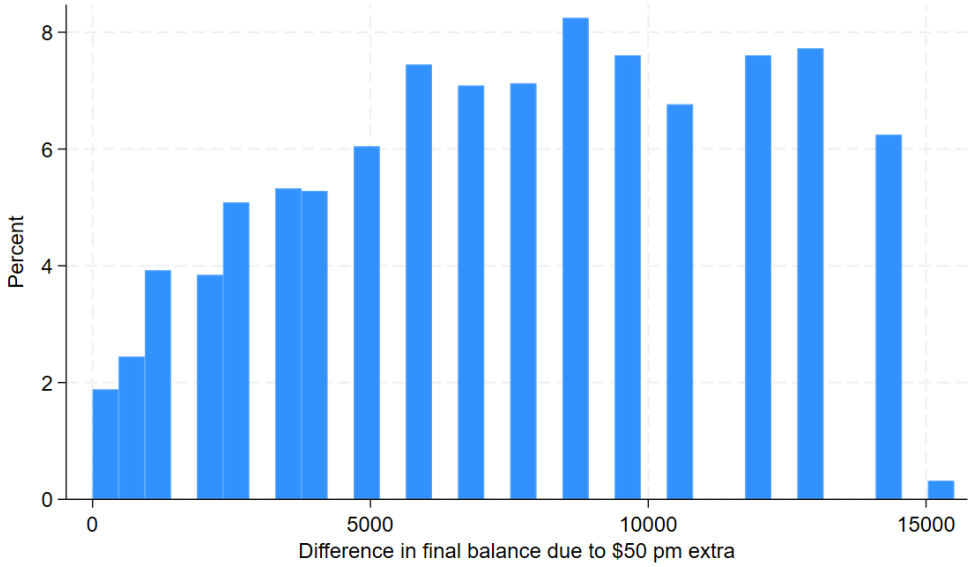
Notes. The figure shows the average amount of monthly contributions into 529 plans, twelve months before and twelve months after opting into an automatic contribution (i.e., a self-selected default). The error bars show a 95% confidence interval around each estimate. The vertical red line represent the month of setting up the default. The analysis includes 118,522 accounts that had data for at least 30 days before and after setting up a default.

Figure 7: The long-term effects of defaults on 529 plans contributions



Notes. The figure shows how cumulative total contributions over an account's lifetime tends to vary based on the age of the account beneficiary at the time of account opening. The 160,517 accounts included in this visual are those with a beneficiary between the ages of 0 and 18 at the time of opening, were closed by the end of 2020, and for which we have transaction history and data on default behavior. Accounts using dynamic default are defined as those that actively end their default at some point before the day of account closure while static default are those that do not end their default until the day of account closure. Sample sizes are respectively: DD: 8,931; SD: 8,820; ND: 142,766.

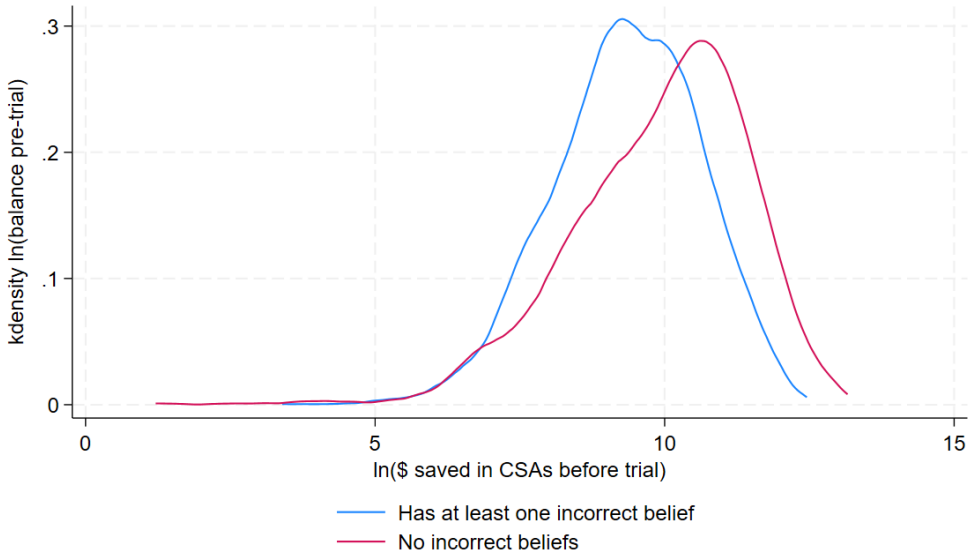
Figure 8: Underestimating the power of small contribution increases



Notes. The figure shows the amount of forgone savings from not increasing one's monthly contributions by $50 per month by the time the account beneficiary reaches college age, for the subset of survey respondents who believed that such an increase would not make a difference.

9 Appendix B: Biased Beliefs and Savings Behavior

Figure 9: Savings balance distribution



Notes. The graph shows the k-density distribution of the ln(balance) variable across owned savings accounts, by biased belief status.

Table 2: Beliefs correlation

	Inertia	Norm misperceptions	Underest. future costs of college	Underest. power of small increases in contributions
Inertia	1			
Norm misperceptions	0.056	1		
Underest. future costs of college	-0.012	0.063	1	
Underest. Power of small increases in contributions	0.017	-0.007	0.052	1

Notes. The table shows the correlation across behavioral barriers.

Table 3: Correlates of Savings Balance (Pre-Trial)

Balance pre-trial	(1)	(2)	(3)
Inertia	-5,557***	-6,338***	-7,661***
	(1,237)	(1,152)	(1,106)
Underest. future costs of college	-15,950***	-17,155***	-12,827***
	(1,266)	(1,305)	(1,271)
Underest. small incr. effect by 10%	-22,703***	-17,508***	-16,934***
	(2,063)	(1,925)	(1,850)
Misperc. saving norms	-12,713***	-14,094***	-11,425***
	(1,835)	(1,725)	(1,658)
HH inc $110K-$159K		4,022**	5,354***
		(1,701)	(1,631)
HH inc $160K-$240K		9,472***	10,098***
		(1,688)	(1,618)
HH inc >$240K		27,444***	28,787***
		(1,701)	(1,630)
Has Financial Advisor		-1,627	-2,657**
		(1,166)	(1,119)
Female		-8,173***	-9,311***
		(1,218)	(1,169)
Owns 2 accounts		97.06	-2,244*
		(1,320)	(1,273)
Owns 3 or more accounts		-7,937***	-13,752***
		(1,544)	(1,520)
White		-1,060	257.2
		(1,339)	(1,285)
Expects 4-yr private coll.		26,836***	22,974***
		(2,206)	(2,128)
Expects 4-yr public coll.		3,299**	1,296
		(1,327)	(1,276)
Child age 6-9			10,377***
			(1,557)
Child age 10-13			20,470***
			(1,589)
Child age 14-16			25,731***
			(1,761)
Child age 17-18			31,696***
			(2,246)
Constant	41,556***	34,262***	21,082***
	(1,035)	(1,936)	(2,047)
N	3,825	3,825	3,825
R^2	0.092	0.227	0.292

Notes. OLS regressions. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Column (1) reports the coefficient of each behavioral barrier dummy on savings balance pre-trial. Column (2) replicates the regression including the available socio-economic indicators. Column (3) includes child age fixed effects.

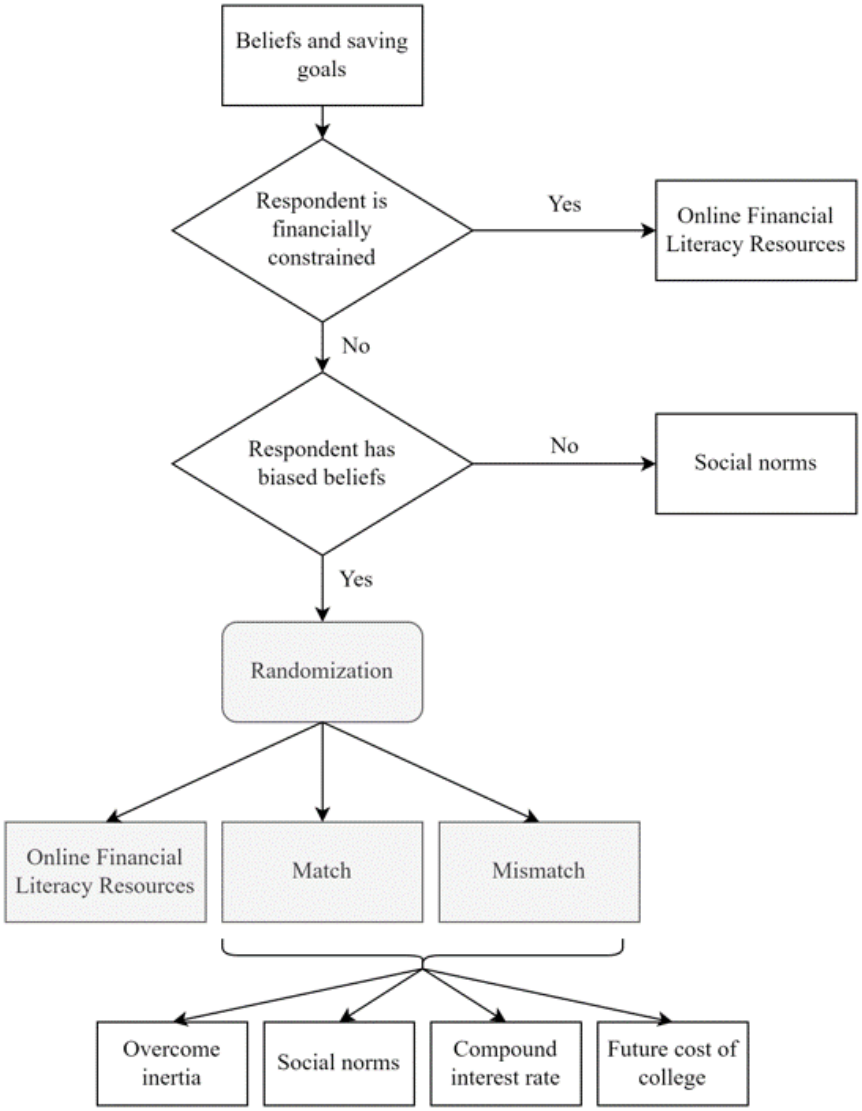
Table 4: Correlates of Each Biased Belief

	(1)	(2)	(3)	(4)
	Inertia	*Misp. norms*	*Under. costs*	*Under. small contr.*
HH inc $110K-$159K	0.0722***	0.0564***	-0.0749***	-0.0238*
	(0.0238)	(0.0159)	(0.021)	(0.0142)
HH inc $160K-$240K	0.0980***	0.0789***	-0.0792***	-0.0401***
	(0.0236)	(0.0157)	(0.0208)	(0.0141)
HH inc >$240K	0.0740***	0.0691***	-0.188***	-0.0770***
	(0.0235)	(0.0157)	(0.0207)	(0.014)
Has Financial Advisor	-0.0604***	0.00987	-0.000782	0.0118
	(0.0164)	(0.011)	(0.0144)	(0.00979)
Female	-0.00907	-0.0340***	0.0343**	0.0232**
	(0.0171)	(0.0115)	(0.0151)	(0.0102)
Owns 2 Accounts	0.0589***	0.0473***	0.00254	0.00885
	(0.0186)	(0.0125)	(0.0164)	(0.0111)
Owns 3 Accounts	0.0284	0.105***	-0.00514	0.0550***
	(0.0221)	(0.0148)	(0.0195)	(0.0132)
White	-0.0211	0.0366***	0.0255	-0.00349
	(0.0188)	(0.0126)	(0.0166)	(0.0112)
Expects 4-yr private coll.	-0.0469	-0.0126	0.537***	-0.00566
	(0.0295)	(0.0197)	(0.026)	(0.0176)
Expects 4-yr public coll.	-0.0273	-0.00811	0.366***	0.0261**
	(0.0174)	(0.0116)	(0.0153)	(0.0104)
Child age 6-9	0.0463**	-0.0238	-0.0989***	-0.00816
	(0.0228)	(0.0152)	(0.0201)	(0.0136)
Child age 10-13	0.0564**	-0.0696***	-0.171***	0.0214
	(0.0231)	(0.0154)	(0.0203)	(0.0138)
Child age 14-16	0.105***	-0.0721***	-0.216***	0.00451
	(0.0254)	(0.017)	(0.0224)	(0.0152)
Child age 17-18	0.0471	-0.103***	-0.244***	-0.0796***
	(0.0325)	(0.0217)	(0.0286)	(0.0194)
Constant	0.414***	0.0666***	0.416***	0.0989***
	(0.0282)	(0.0189)	(0.0249)	(0.0169)
N	3,825	3,825	3,825	3,825
R^2	0.021	0.04	0.215	0.027
Avg. Dep. Var.	*0.483*	*0.131*	*0.412*	*0.0996*

Notes. OLS regressions. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The first row reports which behavioral barrier dummy is included as the outcome variable, regressed on the complete set of available socio-economic indicators and child age fixed effects.

10 Appendix C: Experimental Design and Results

Figure 10: Experimental survey flow



Notes. The image shows the flow of the main experimental survey and the inclusion criteria in the randomization process.

10.1 Treatment messages

Control: Online self-help resources. *We know how important it is to save for college. To help you save better, we have created College Savings U, an online resource with videos, college saving calculators, webinars, and much else to help you understand the cost of college, how far your savings might go, and how to make the best use of your 529 College Savings Plan. You access these resources at any time here.* The word "here" contained a hyperlink redirecting respondents to the following webpage: `https://brightstart.com/be-ready/`

Adjusting defaults. *Automatic transfers (called an Automatic Investment Plan, or AIP) can help you avoid forgetting to contribute regularly and help you reach your saving goals. However, it is a good idea to review the amount and frequency of your automatic transfers to avoid anchoring yourself to the same amounts over a long time. A small increase in the amount of your 529 recurring savings today can make a big difference tomorrow. Don't delay such an important decision. Review now the amount and frequency of your 529 contributions to make sure you reach your saving goals. You can login into your account by clicking here.*

Saving norms. *Most people underestimate how much they should save for college, and whether their current 529 savings are in line with other account owners. The table below can help you understand where your current saving lies in comparison to other account owners with beneficiaries of similar age as yours. A small amount increase in your 529 regular savings today can make a big difference tomorrow. Don't delay such an important decision. Review the amount and frequency of your 529 contributions today to make sure you reach your saving goals. You can login into your account by clicking here.*

Beneficiary age	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile
6	\$1,480	\$4,190	\$11,288	\$28,294	\$59,155
7	\$1,646	\$4,606	\$12,024	\$30,817	\$65,480
8	\$1,752	\$5,041	\$13,490	\$33,361	\$69,708
9	\$1,663	\$5,189	\$14,264	\$37,429	\$74,157

Notes. The table was age-adjusted based on the respondent's beneficiary age group (1-5, 6-9, 10-13, and 14-17). Here only the 6-9 group is shown for brevity.

Future cost of college. *Most people underestimate how much they should save for college. The table below shows the forecast total cost of college for beneficiaries of similar age as yours based on cost trends over the past several decades.*

Current Age	Year of Enrollment	Public 4-Year (In-State)	Public 4-Year (Out-of-State)	Private 4-Year
6	2034	$185,228	$315,421	$409,698
7	2033	$176,408	$300,401	$390,189
8	2032	$168,007	$286,096	$371,608
9	2031	$160,007	$272,472	$353,913

Notes. The table was age-adjusted based on the respondent's beneficiary age group (1-5, 6-9, 10-13, and 14-17). Here only the 6-9 group is shown for brevity.

The power of small increases in contributions. *Most people underestimate the power of small increases in savings activity. A small increase today to the amount of your 529 regular savings can make a big difference tomorrow. The table below shows that modest amounts of monthly saving can have a big impact over time when accounting for a 5% annual return. Don't delay such an important decision, and review the amount and frequency of your 529 contributions today to make sure you can reach your saving goals. You can login into your account by clicking here.*

		Years of saving			
		4 Years	8 Years	12 Years	16 Years
Monthly contribution	$50	$2,700	$6,000	$10,000	$14,900
	$100	$5,400	$12,000	$20,000	$29,800
	$150	$8,100	$18,000	$30,000	$44,700
	$200	$10,900	$24,100	$40,100	$59,600
	$250	$13,600	$30,100	$50,100	$74,000
	$300	$16,300	$36,100	$60,200	$89,200

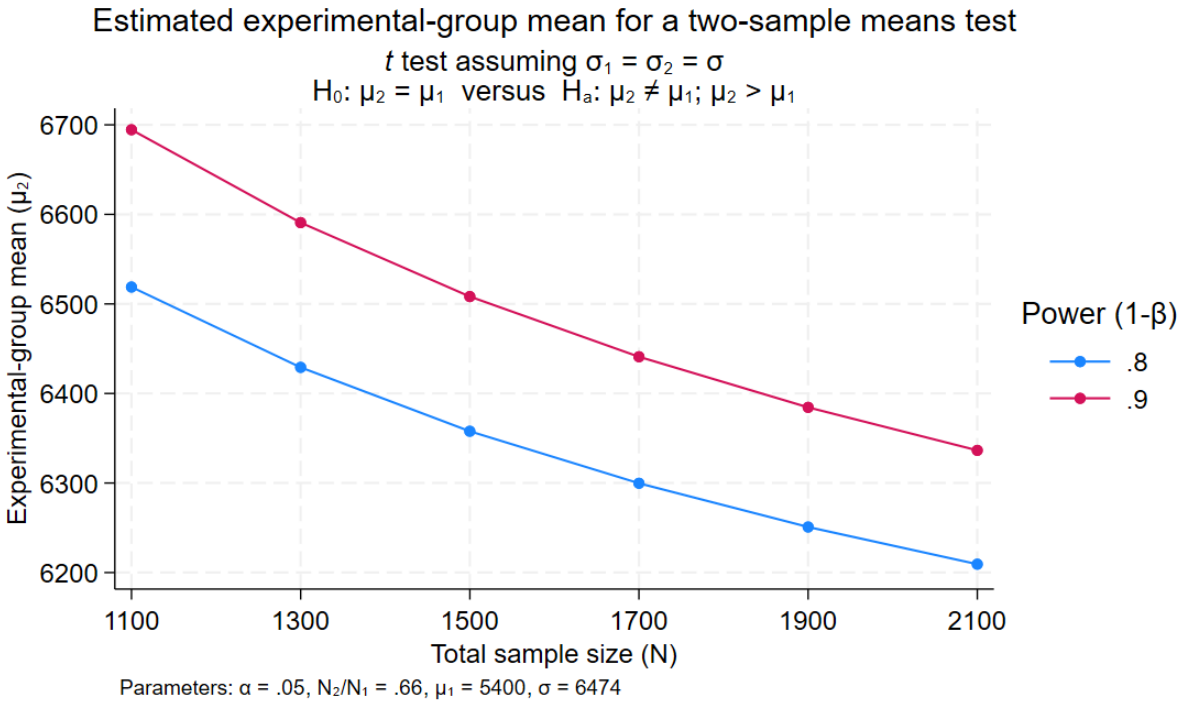
10.2 Power analysis and balance checks

Table 5: Balance checks: Control vs Pooled Treatment

	Control			Treatment			
	N	mean	sd	N	mean	sd	Diff
Balance pre-trial	450	25284	36666	1784	25615	39064	331
HH income	450	195094	147042	1784	209260	247522	14,166
Has Financial Advisor	450	0.44	0.50	1784	0.49	0.50	0.051*
Female	450	0.39	0.49	1784	0.37	0.48	-0.022
Owns more than 1 account	450	0.62	0.49	1784	0.65	0.48	0.029
Is main contributor	450	0.96	0.19	1784	0.95	0.21	-0.010
Uses automatic transfers	450	0.65	0.48	1784	0.62	0.49	-0.029
White	450	0.73	0.45	1784	0.75	0.43	0.022
Expects 4-yr private coll.	450	0.09	0.28	1784	0.07	0.26	-0.012
Eldest beneficiary's age	450	9.88	4.82	1784	9.95	4.86	0.077
Respondent's age	450	48.61	12.42	1784	48.00	12.72	-0.613

Notes. The table reports the summary staistics for the randomized control and pooled treatment groups, and their relative differences. *** p<0.01, ** p<0.05, * p<0.1.

Figure 11: Power calculations



The figure shows the required experimental group mean given different sample sizes. The control group mean of annual contributions is \$5,400 (with s.d. of \$6,474) based on administrative data.

10.3 Experimental results

Note: the controls specified at the bottom of each regression table are the following: Balance pre-trial is the savings balanced saved up until one month (the latest pre-trial data pull) prior to the survey participation; Behavioral barriers is a categorical variable indicating which of the four selected barriers the individual displayed in the survey; socio-economic characteristics include a dummy indicating if the respondent is a parent (with the alternative being a grandparent or close relative responsible for financing the beneficiary's college education), a dummy indicating if they report usually relying on a financial planner or accountant for saving decisions, the number of beneficiaries they need to fund college for, the number of 529 plans currently owned and open, if they are the main contributors to the 529 plan, account

owner race, account owner gender, and the college type they hope their beneficiary will attend (4-year public in-state, 4-year public out-of-state, 4-year private, community college, apprenticeship). Child age group fixed effects is a categorical variable that aims to control for the beneficiary's proximity to college age, thus capturing the marginally decreasing returns from additional contributions.

Table 6: Treatment effects of nudge messages on savings

DV: Δ *Saving Balance*	(1)	(2)	(3)	(4)
Procrastinate msg	0.289***	0.221***	0.174***	0.158**
	(0.0618)	(0.0639)	(0.0619)	(0.0619)
Norms msg	-0.0951*	-0.0980*	-0.121**	-0.124**
	(0.0572)	(0.0576)	(0.0533)	(0.0537)
Compound msg	0.164***	0.193***	0.174***	0.175***
	(0.0593)	(0.0594)	(0.0573)	(0.0567)
Futre costs msg	-0.057	-0.0871	-0.0324	-0.0344
	(0.0635)	(0.0651)	(0.0619)	(0.0618)
Avg. Dep. Var.	*5,548*	*5,548*	*5,548*	*5,548*
Balance pre-trial	Yes	Yes	Yes	Yes
Beliefs		Yes	Yes	Yes
Socio-economic controls			Yes	Yes
Child age group F.E.				Yes
Constant	7.811***	7.832***	7.694***	7.734***
	(0.0532)	(0.0672)	(0.0815)	(0.0826)
N	2,234	2,234	2,234	2,234

Notes. GLM regressions. Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1. The table shows the results of a regression estimating the effects of each nudge message on savings balance differences between the period pre-trial and one year after survey completion. Each column reports the coefficient estimates with and without the inclusion of controls.

Table 7: Pooled treatments

Δ *Balance after 1 year*	(1)	(2)	(3)
Relevant	0.106*	0.076	0.0694
	(0.0544)	(0.0517)	(0.0514)
Irrelevant	-0.0006	-0.0517	-0.0052
	(0.0504)	(0.0474)	(0.0473)
Avg. Dep. Var.	*5,548*	*5,548*	*5,548*
Balance pre-trial	Yes	Yes	Yes
Socio-economic controls		Yes	Yes
Child age group F.E.			Yes
Constant	7.775***	7.599***	7.614***
	(0.001)	(0.0075)	(0.769)
N	2,234	2,234	2,234

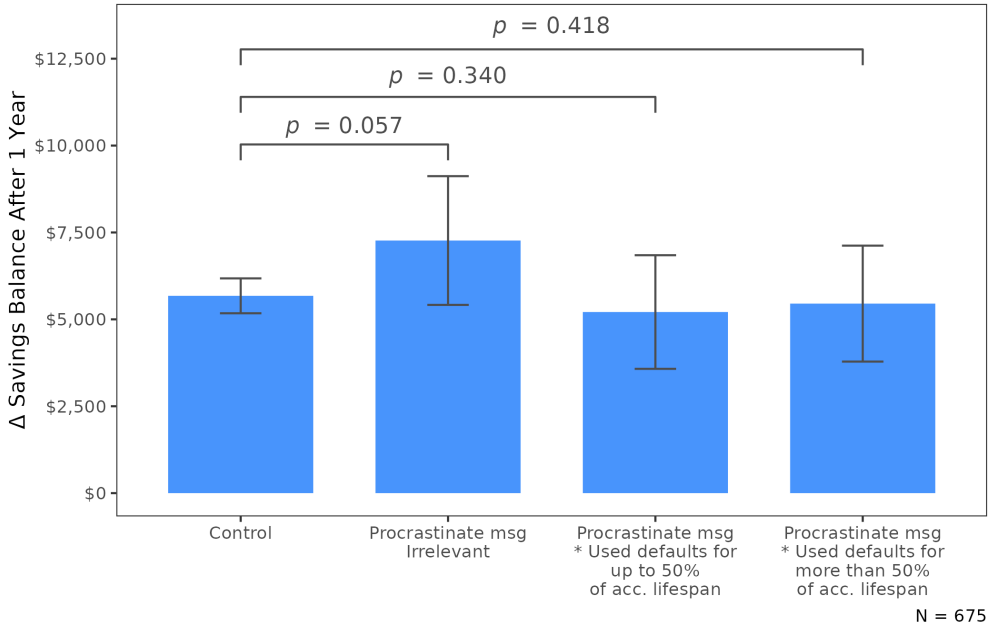
Notes. GLM regressions. Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1. Columns (1) to (3) report the coefficients of the pooled nudge messages defined as relevant or irrelevant by matching the behavioral barrier to the nudge message saw by the respondent.

Table 8: Made a deposit within 30 days post-treatment

DV: *Made a deposit*	(1)	(2)	(3)
Relevant	0.162*	0.164*	0.174**
	(0.0833)	(0.0842)	(0.0842)
Irrelevant	-0.140*	-0.141*	-0.141*
	(0.0725)	(0.0733)	(0.0734)
Avg. dep. var.	0.677	0.677	0.677
Balance pre-trial	Yes	Yes	Yes
Socio-economic controls		Yes	Yes
Child age group FE			Yes
Constant	0.577***	0.508***	0.480***
	(0.0649)	(0.108)	(0.115)
N	2,234	2,234	2,234

Notes. GLM regressions. Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1. Columns (1) to (3) report the coefficients of the pooled nudge messages defined as relevant or irrelevant by matching the behavioral barrier to the nudge message saw by the respondent.

Figure 12: Relevance by barrier type: Overcoming inertia



Notes. The graph shows the average difference, and the 95% confidence interval, in savings balance between the month before survey completion and 12 months after for the control group, and the interaction term of the procrastinate message and the categorical variable indicating the extent to which a person utilized defaults.

Table 9: Inertia vs. Control

DV: Δ *Saving Balance*	(1)	(2)	(3)
Procrastinate msg	0.208***	0.247*	0.247*
	(0.0618)	(0.130)	(0.130)
Inertia	0.0625	0.0755	0.205*
	(0.0662)	(0.0771)	(0.117)
Procrastinate msg * Inertia		-0.0543	
		(0.145)	
Used AIP for up to 50%			0.0898
			(0.0941)
Used AIP for more than 50%			0.0697
			(0.0861)
Procrastinate msg * Used AIP for up to 50%			-0.0856
			(0.160)
Procrastinate msg * Used AIP for more than 50%			-0.0402
			(0.156)
Average Dep. Var.	5,677	5,677	5,677
Balance pre-trial	Yes	Yes	Yes
Beliefs	Yes	Yes	Yes
Socio-economic controls	Yes	Yes	Yes
Child age group F.E.	Yes	Yes	Yes
Constant	7.744***	7.735***	7.737***
	(0.124)	(0.128)	(0.129)
N	675	675	675

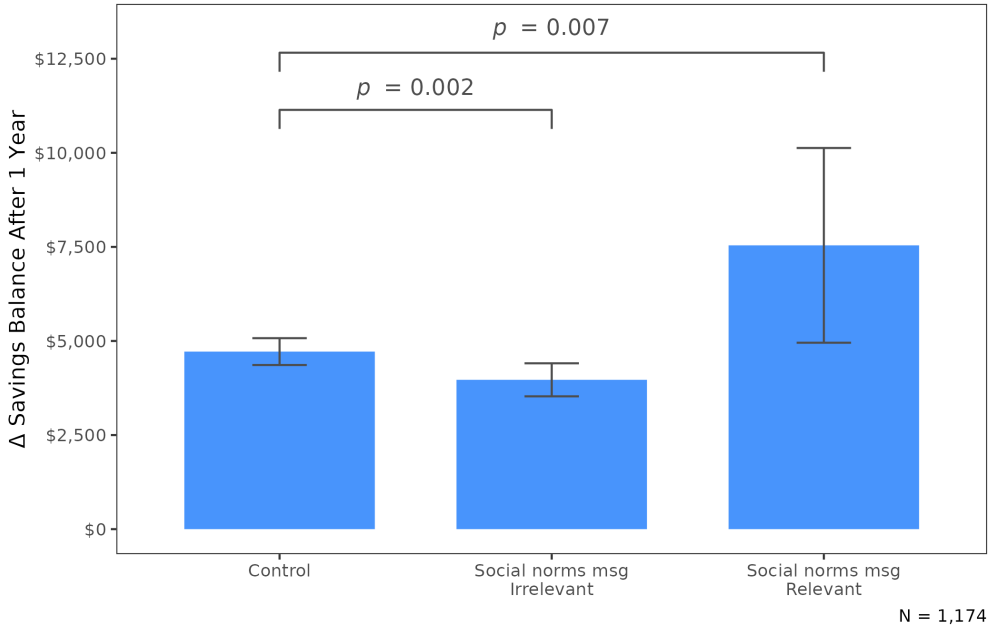
Notes. GLM regressions. Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1. Column (1) reports the coefficients of the procrastinate message and the related behavioral barrier. Column (2) reports the interaction between the message and the barrier. Column (3) reports the interaction between the message and the barrier split up by default usage percentage.

Table 10: Inertia vs. Irrelevant

DV: Δ *Saving Balance*	(1)	(2)
Procrastinate msg	0.164**	0.162*
	(0.0681)	(0.0976)
Norms msg	-0.102	-0.0583
	(0.0628)	(0.102)
Compound msg	0.0742	0.103
	(0.0708)	(0.104)
Future costs msg	-0.00895	0.273*
	(0.0744)	(0.146)
Used AIP for more than 50%	-0.0786*	0.000158
	(0.0467)	(0.0932)
Procrastinate msg * Used AIP for more than 50%		0.00592
		(0.130)
Norms msg * Used AIP for more than 50%		-0.0604
		(0.128)
Compound msg * Used AIP for more than 50%		-0.0334
		(0.135)
Future costs msg * Used AIP for more than 50%		-0.379**
		(0.169)
Avg. Dep. Var.	5,501	5,501
Balance pre-trial	Yes	Yes
Beliefs	Yes	Yes
Socio-economic controls	Yes	Yes
Child age group F.E.	Yes	Yes
Constant	7.876***	7.806***
	(0.105)	(0.112)
N	1,152	1,152

Notes. GLM regressions. Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1. Column (1) reports the coefficients of the norms message and the related behavioral barrier. Column (2) reports the interaction between the message and the barrier. Column (3) reports the interaction between the message and the barrier split up by tercile of the distribution of the norm misperception.

Figure 13: Relevance by barrier type: Norm misperceptions



Notes. The graph shows the average difference, and the 95% confidence interval, in savings balance between the month before survey completion and 12 months after for the control group, and the interaction term of the social norms message and the dummies flagging whether the respondent misperceived the saving norm (relevant) or not (irrelevant).

Table 11: Norms vs. Control

DV: Δ *Saving Balance*	(1)	(2)	(3)
Norms msg	-0.122**	-0.173***	-0.0274
	(0.0540)	(0.0564)	(0.0926)
Misperc. saving norms	0.257***	-0.0461	0.444*
	(0.0915)	(0.140)	(0.247)
Norms msg * Misperc. saving norms		0.469***	
		(0.175)	
Saves up to $14K less than peers			-0.482***
			(0.115)
Saves $14-24K less than peers			-0.966***
			(0.120)
Saves $24-47K less than peers			-1.356***
			(0.144)
Norms msg * Saves up to $14K less than peers			0.0577
			(0.133)
Norms msg * Saves $14-24K less than peers			-0.0349
			(0.133)
Norms msg * Saves $24-47K less than peers			-0.113
			(0.149)
Mean dep. Var	4,718	4,718	4,718
Behavioral barriers	Yes	Yes	Yes
Balance pre-trial	Yes	Yes	Yes
Socio-economic controls	Yes	Yes	Yes
Child age group F.E.	Yes	Yes	Yes
Constant	7.514***	7.546***	8.243***
	(0.103)	(0.104)	(0.129)
N	1,174	1,174	1,174

Notes. GLM regressions. Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1. Column (1) reports the coefficients of the norms message and the related behavioral barrier. Column (2) reports the interaction between the message and the barrier. Column (3) reports the interaction between the message and the barrier split up by tercile of the distribution of the norm misperception.

Table 12: Norms vs. Irrelevant

DV: Δ *Saving Balance*	(1)	(2)
Procrastinate msg	0.351**	0.329*
	(0.141)	(0.193)
Norms msg	0.294**	0.249
	(0.120)	(0.153)
Compound msg	0.263**	0.141
	(0.132)	(0.152)
Futre costs msg	0.198	0.198
	(0.124)	(0.160)
Saves $14-24K less than peers		-0.392
		(0.291)
Saves $24-47K less than peers		0.0504
		(0.259)
Procrastinate msg * Saves $14-24K less than peers		0.0743
		(0.367)
Procrastinate msg * Saves $24-47K less than peers		0.313
		(0.33)
Norms msg * Saves $14-24K less than peers		0.288
		(0.327)
Norms msg * Saves $24-47K less than peers		0.13
		(0.336)
Compound msg * Saves $14-24K less than peers		0.368
		(0.408)
Compound msg * Saves $24-47K less than peers		0.324
		(0.341)
Future costs msg * Saves $14-24K less than peers		0.0489
		(0.323)
Future costs msg * Saves $24-47K less than peers		0.134
		(0.306)
Mean dep. Var	4,718	4,718
Beliefs	Yes	Yes
Balance pre-trial	Yes	Yes
Socio-economic controls	Yes	Yes
Child age group F.E.	Yes	Yes
Constant	7.514***	7.546***
	(0.103)	(0.104)
N	1,174	1,174

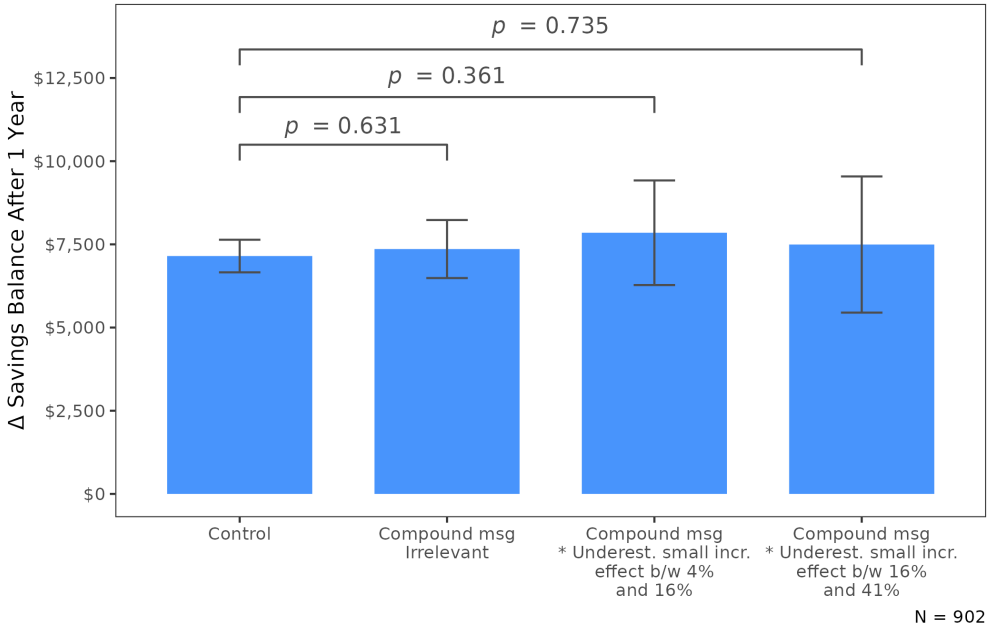
Notes. GLM regressions. Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1. Column (1) reports the coefficients of each nudge message, while Column (2) reports the interactions between the categorical variable of the behavioral barrier with each of the nudge messages.

Table 13: Compound vs. Control Message

DV: Δ *Saving Balance*	(1)	(2)	(3)
Compound msg	0.258***	0.361***	0.0290
	(0.0546)	(0.0552)	(0.0604)
Underest. small incr. effect by 10%	-0.664***	-0.340**	
	(0.0919)	(0.148)	
Compound msg * Underest. small incr. effect by 10%		-0.593***	
		(0.171)	
Underest. small incr. effect b/w 4% and 16%			-0.679***
			(0.0899)
Underest. small incr. effect b/w 16% and 41%			-1.598***
			(0.108)
Compound msg * Underest. small incr. effect b/w 4% and 16%			0.0934
			(0.102)
Compound msg * Underest. small incr. effect b/w 16% and 41%			0.0472
			(0.139)
Avg. Dep. Var.	7149	7149	7149
Balance pre-trial	Yes	Yes	Yes
Beliefs	Yes	Yes	Yes
Socio-economic controls	Yes	Yes	Yes
Child age group F.E.	Yes	Yes	Yes
Constant	8.108***	8.080***	8.957***
	(0.122)	(0.118)	(0.121)
N	902	902	902

Notes. GLM regressions. Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1. Column (1) reports the coefficients the relevant nudge message and the behavioral barrier. Column (2) includes the interaction effect between these two terms. Column (3) reports the coefficients of the interaction terms between the nudge message and the categorical variable of the behavioral barrier.

Figure 14: Relevance by barrier type: Underestimating compound interest effects



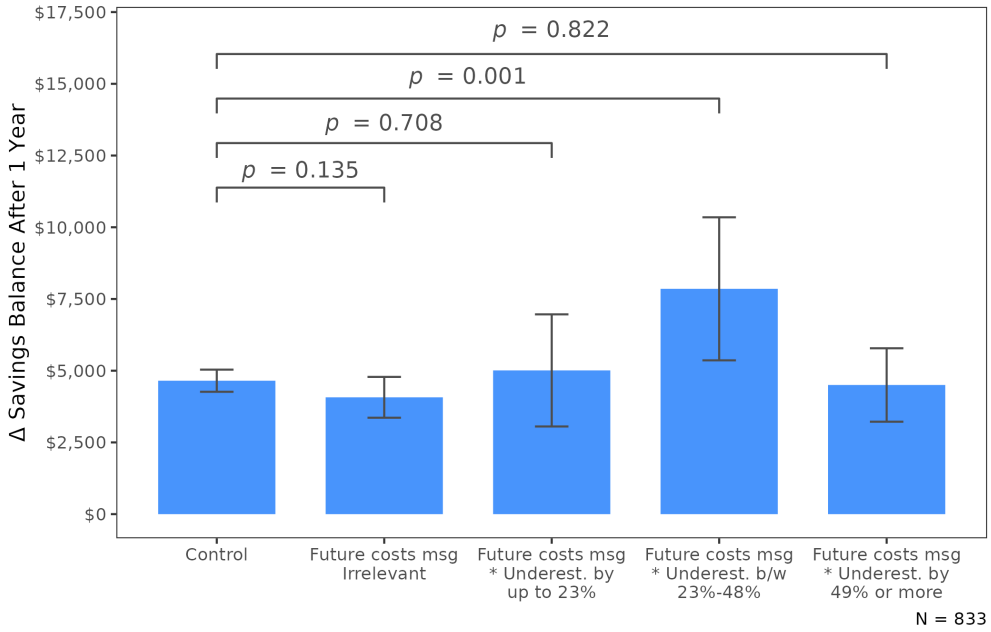
Notes. The graph shows the average difference, and the 95% confidence interval, in savings balance between the month before survey completion and 12 months after for the control group, and the interaction term of the compound interest message and the categorical variable indicating the extent to which a person underestimated the value of small contributions thanks to compound interest effects.

Table 14: Compound vs. Irrelevant

DV: Δ *Saving Balance*	(1)	(2)
Procrastinate msg	0.418	0.636*
	(0.268)	(0.330)
Norms msg	0.0854	0.0741
	(0.148)	(0.266)
Compound msg	-0.0621	-0.161
	(0.146)	(0.244)
Future costs msg	0.0409	0.437
	(0.171)	(0.303)
Underest. small incr. effect b/w 16% and 41%		0.110
		(0.284)
Procrastinate msg * Underest. small incr. effect b/w 16% and 41%		-1.134***
		(0.383)
Norms msg * Underest. small incr. effect b/w 16% and 41%		0.00865
		(0.314)
Compound msg * Underest. small incr. effect b/w 16% and 41%		0.140
		(0.293)
Future costs msg * Underest. small incr. effect b/w 16% and 41%		-0.552
		(0.364)
Avg. Dep. Var.	2,285	2,285
Balance pre-trial	Yes	Yes
Beliefs	Yes	Yes
Socio-economic controls	Yes	Yes
Child age group F.E.	Yes	Yes
Constant	6.984***	6.910***
	(0.219)	(0.333)
N	301	301

Notes. GLM regressions. Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1. Column (1) reports the coefficients of each nudge message. Column (2) reports the coefficients of the interactions between each nudge message and the behavioral barrier.

Figure 15: Relevance by barrier type: Underestimating future costs



Notes. The graph shows the average difference, and the 95% confidence interval, in savings balance between the month before survey completion and 12 months after for the control group and the interaction between the future cost of college message and the categorical variable measuring the extent to which respondents nderestimated the future cost of college.

Table 15: Costs vs. Control

DV: Δ *Saving Balance*	(1)	(2)	(3)
Future costs msg	-0.0383	-0.100	-0.102
	(0.0602)	(0.0894)	(0.0894)
Underest. future costs of college	0.111	0.0676	
	(0.0682)	(0.0855)	
Future costs msg * Underest. future costs		0.107	
		(0.119)	
Underest. by up to 23%			0.210*
			(0.115)
Underest. b/w 24%-48%			-0.198
			(0.123)
Underest. by 49% or more			0.0898
			(0.112)
Future costs msg * Underest. by up to 23%			0.0683
			(0.191)
Future costs msg * Underest. b/w 23%-48%			0.493***
			(0.163)
Futre costs msg * Underest. by 49% or more			-0.0578
			(0.145)
Avg. Dep Var.	*4,651*	*4,651*	*4,651*
Balance pre-trial	Yes	Yes	Yes
Beliefs	Yes	Yes	Yes
Socio-economic controls	Yes	Yes	Yes
Child age group F.E.	Yes	Yes	Yes
Constant	7.553***	7.582***	7.614***
	(0.119)	(0.127)	(0.128)
N	833	833	826

Notes. GLM regressions. Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1. Column (1) reports the coefficients of the nudge message and the related behavioral barrier. Column (2) reports the coefficient of the interaction between these two terms. Column (3) reports the coefficients of the interactions between the nudge message and the categorical variable of the behavioral barrier.

Table 16: Costs vs. Irrelevant

DV: Δ *Saving Balance*	(1)	(2)
Procrastinate msg	0.161*	0.0262
	(0.0954)	(0.125)
Norms msg	-0.148*	-0.484***
	(0.0780)	(0.131)
Compound msg	0.157	0.132
	(0.0983)	(0.135)
Future costs msg	0.00505	-0.00645
	(0.0787)	(0.171)
Underest. b/w 23%-48%		-0.396***
		(0.141)
Underest. by 49% or more		-0.105
		(0.130)
Procrastinate msg * Underest. b/w 23%-48%		0.419*
		(0.232)
Procrastinate msg * Underest. by 49% or more		-0.00833
		(0.232)
Norms msg * Underest. b/w 23%-48%		0.671***
		(0.184)
Norms msg * Underest. by 49%		0.358**
		(0.179)
Compound msg * Underest. b/w 23%-48%		-0.0268
		(0.206)
Compound msg * Underest. by 49%		0.0520
		(0.275)
Future costs msg * Underest. b/w 23%-48%		0.406*
		(0.215)
Future costs msg * Underest. by 49%		-0.174
		(0.204)
Avg. Dep. Var.	*4,322*	*4,322*
Balance pre-trial	Yes	Yes
Beliefs	Yes	Yes
Socio-economic controls	Yes	Yes
Child age group F.E.	Yes	Yes
Constant	7.607***	7.768***
	(0.114)	(0.139)
N	961	961

Notes. GLM regressions. Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1. Column (1) reports the coefficients of each nudge message. Column (2) reports the coefficients of the interactions between each nudge message and the behavioral barrier.

Table 17: Time spent on the survey

DV: *Minutes spent to complete the survey*	(1)	(2)	(3)
Relevant	-0.0633	0.191	0.550
	(0.476)	(0.427)	(0.370)
Irrelevant	-0.163	-0.140	0.0294
	(0.457)	(0.380)	(0.313)
Avg. dep. var.	4225	4225	4225
Balance pre-trial	Yes	Yes	Yes
Socio-economic controls		Yes	Yes
Child Age Group FE			Yes
Constant	8.537***	8.644***	7.722***
	(0.394)	(0.540)	(0.507)
N	2,234	2,234	2,234

Notes. GLM regressions. Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1. Columns (1) to (3) report the coefficients of the pooled nudge messages defined as relevant or irrelevant by matching the behavioral barrier to the nudge message saw by the respondent.

Table 18: Lottery allocation preferences

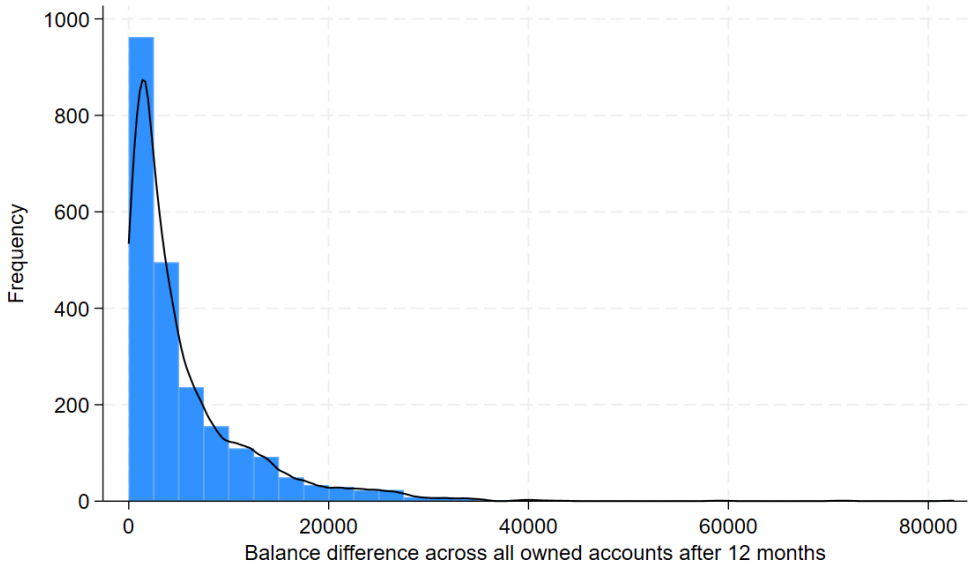
DV: *Amount of $150 lottery on deposit*	(1)	(2)	(3)
Relevant	30.16	36.77	34.79
	(44.25)	(44.06)	(44.14)
Irrelevant	57.65	56.02	57.16
	(39.58)	(39.27)	(39.33)
var(e.lottery)	310,120***	299,760***	299,383***
	(43,030)	(41,566)	(41,512)
Avg. Dep. Var.	111.5	111.5	111.5
Balance pre-trial	Yes	Yes	Yes
Socio-economic controls		Yes	Yes
Child age group F.E.			Yes
Constant	446.1***	486.7***	506.7***
	(42.07)	(63.58)	(67.72)
N	2,234	2,234	2,234

Notes. GLM regressions. Robust standard errors in parentheses. ***p<0.01, **p<0.05, *p<0.1. Columns (1) to (3) report the coefficients of the pooled nudge messages defined as relevant or irrelevant by matching the behavioral barrier to the nudge message saw by the respondent.

11 Appendix E: Empirical strategy

In this Appendix we provide evidence on the rationale behind the choice of our empirical strategy. We begin by plotting the distribution of our outcome variable: saving balances across owned accounts after a year from completing the survey. From Figure 16 we can already see the variable is right-skewed with a long tail distribution, meaning that a small share of account owners in our sample saved very large sums of money over the period of one year. This should be expected, since we knew from pre-trial analysis on administrative data that there are large savings inequalities among account owners.

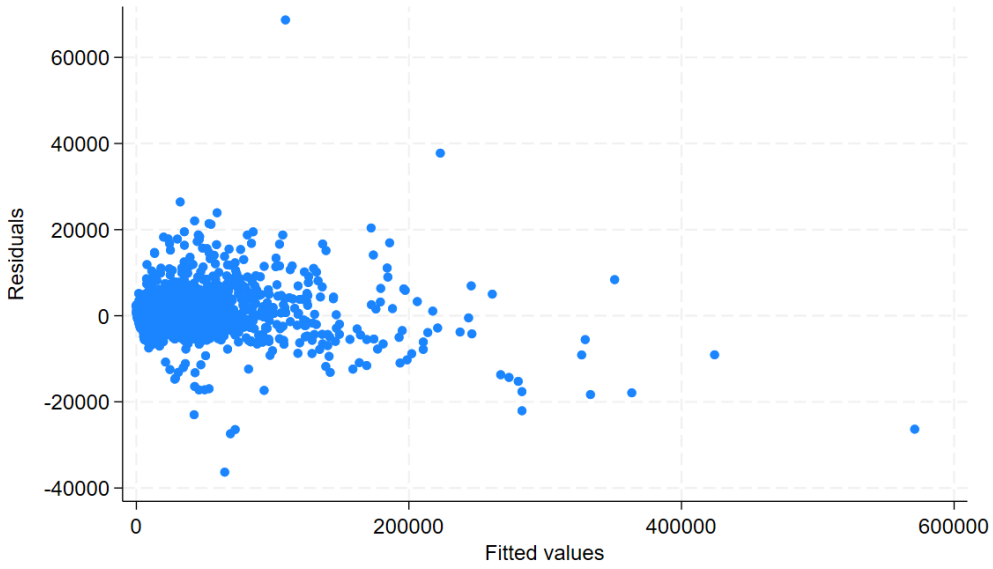
Figure 16: Outcome variable distribution



Notes. The figure shows the distribution of the outcome variable.

We then the same regression as in Equation 1 but using a linear probability model. An important assumption of this model is that the variance in the residuals has to be homoskedastic or constant (Stock and Watson, 2008).

Figure 17: OLS residuals plot



Notes. The figure shows the distribution of the residuals after the linear probability regression using Stata's *rvfplot* command.

A Shapiro-Wilk test for normality confirms that we reject the null hypothesis that the distribution of the residuals is normal (W=0.788, V=277.553, z=14.369, p<0.000).

Table 19: Variance inflation factor

Variables	VIF	1/VIF
Balance pre-trial	2.00	0.500195
Inertia	2.00	0.500663
Underest. future costs of college	2.22	0.45016
Underest. small incr. effect by 10%	1.20	0.834423
Misperc. saving norms	1.22	0.820095
HH inc $110K-$159K	1.69	0.592478
HH inc $160K-$240K	2.35	0.425278
HH inc >$240K	2.17	0.460897
Has Financial Advisor	1.99	0.501635
Female	1.66	0.60402
2 Accounts Open	2.11	0.473293
3 Accounts Open	1.86	0.538013
White	3.55	0.282007
Expects 4-yr private coll.	1.38	0.726695
Expects 4-yr public coll.	1.92	0.521053
Child age 6-9	1.75	0.57184
Child age 10-13	1.89	0.528768
Child age 14-16	1.79	0.557332
Child age 17-18	1.50	0.66689
Mean VIF	*1.88*	

Notes. The table reports the results of a Variance Inflation Factor analysis to test for multicollinearity among the independent variables used in the full specification of all regressions reported in the manuscript.

11.0.1 Post-estimation tests

Autocorrelation. In this subsection, we evaluate the validity of our choice for the empirical strategy by means of a post-estimation test. Across all specifications, we always account for pre-trial savings balance by including the lagged dependent variable. This is common practice when testing for changes in outcomes that tend to be relatively static over time, such as savings, but it is still crucial to test for autocorrelation in the residuals because it could indicate that past changes in outcome are influencing current changes, which violates the independence assumption of errors.

To test for the presence of autocrrelation, we run both a post-estimation Durbin-Watson test and a more stringent Breusch-Godfrey test for higher-order autocorrelation.

Table 20: Durbin-Watson test for autocorrelation

	(1)
	Deviance residual
Deviance residual	0.00231
	(0.0212)
Constant	-0.234***
	(0.0185)
N	2,233
R^2	0.000
Standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

The tests yield a Durbin–Watson d-statistic (2, 2233) of 2.000359, suggesting that the residuals of the model are not correlated over time, meaning there is no evidence of either positive or negative autocorrelation. The results of the more stringent Breusch-Godfrey test yield a χ^2=0.006, p=0.9404, confirming that there is no significant evidence of autocorrelation in the residuals of our model, implying that the model does not suffer from serial correlation problems. The absence of autocorrelation suggests that the model's estimates are reliable, and the inference is likely valid.

Endogeneity. In our model 1, testing for endogeneity involves checking if the lagged dependent variable y at time *t-1* is endogenous. We can use an instrumental variable approach to rule out this risk. Of all the covariates we have data for, the type of college that parents expect their child to enrol in can be a valid instrument, in particular whether the survey respondent expects their child to attend a 4-year private University, the most expensive type of higher education institution. Anticipating that a child might attend an expensive college can be related to past expectations and influencing savings so far, making it a relevant instrument. Further, expectations are tied to future intentions rather than immediate changes, minimizing direct impacts on our outcome variable, which is the difference in savings balance within a year from treatment.

To test for endogeneity, we use Wooldridge (2010) approach by including the residuals

from the first-stage regression as an additional regressor in the second stage. In a set of tests from the first stage regression we find that the instrument is valid.

Table 21: IV test for endogeneity

	(1)	(2)
	Δ *saving balance*	Δ *saving balance*
Balance pre-trial	0.113***	0.136***
	(0.00845)	(0.0253)
Expects 4-year Pvt Uni.	603.1	
	(625.3)	
Relevant	301.9	263.0
	(341.1)	(337.4)
Irrelevant	92.81	85.95
	(254.6)	(256.5)
Constant	2,490***	1,960***
	(265.4)	(649.1)
N	2,234	2,234
R^2	0.431	0.414
Kleibergen-Paap rk LM statistic		22.063
Cragg-Donald Wald F statistic		75.553
Kleibergen-Paap rk Wald F statistic		24.737

The tests reported in the table above are supported by an endogeneity test statistic yielding a value of 0.923, p=0.3366, which implies that the difference between the OLS and the IV estimates is not statistically significant, thus ruling out the risk of endogeneity that could have arose from the introduction of our lagged dependent variable across our main specifications.

11.0.2 Data cleaning

In this subsection, we summarize the data cleaning decisions we made before analyzing the trial results. As a first step, we identified five open-text fields within the survey that could be used to flag responses for further manual review: self-reported expected total cost of undergraduate education, self-reported 529 account savings goal, self-reported account balance, self-reported annual household income, and the final pre-treatment question of *'Is*

there anything about this survey you would like to tell us?' Respondents with at least two self-reported numbers flagged had all of their responses examined holistically and compared against administrative data when available before determining a course of action, and those who provided information about themselves indicating that they were unable to answer the survey questions accurately or that they were outside of our population of interest were dropped.

We ultimately removed 32 observations during this process. Removals included those who reported identical values for all questions such as inputting '123456789' as the individual's expected cost of college, current account balance, and household income simultaneously. Additionally, respondents who reported extremely high or extremely low values for all questions were also removed. For example, one respondent reported an expected total cost of college to be $120,000,000, a current account balance of $67,000,000, and a household income of $150,000,000 while multiple respondents reported $0 for all questions. Among the open-text comment space, respondents who indicated an unwillingness to answer ('It is none of your business what my income is or my age!') or an inability to answer ('Some questions only have the ability to answer in numbers, and I didn't know the answer, so I had to make it up') the previous questions were dropped as well. Lastly, some respondents were removed from our data set after proving information about themselves indicating that they did not meet our eligibility criteria such as 'I withdrew and closed the account.' or 'My beneficiary is already in her freshman year of college so many of the questions didn't really apply to my situation'. After excluding responses for any of the above reasons, the final sample size retained was 3,825.

References

Agrawal, K., S. Athey, A. Kanodia, and E. Palikot (2022). Personalized recommendations in edtech: Evidence from a randomized controlled trial. *arXiv preprint arXiv:2208.13940*.

Andrieş, A. M. and S. Walker (2023). When the message hurts: The unintended impacts of nudges on saving. *Journal of Comparative Economics 51*(2), 439–456.

Arni, P., D. Dragone, L. Goette, and N. R. Ziebarth (2021). Biased health perceptions and risky health behaviors—theory and evidence. *Journal of Health Economics 76*, 102425.

Ashraf, N., D. Karlan, and W. Yin (2006). Tying odysseus to the mast: Evidence from a commitment savings product in the philippines. *The Quarterly Journal of Economics 121*(2), 635–672.

Athey, S., S. A. Cole, S. Nath, and S. J. Zhu (2023). Targeting, personalization, and engagement in an agricultural advisory service. *Available at SSRN 4536641*.

Azevedo, V., J. Lafortune, L. Olarte, and J. Tessada (2024). Personalizing or reminding? how to better incentivize savings among underbanked individuals. *Journal of Economic Behavior & Organization 222*, 25–63.

Barber, B. M. and T. Odean (2001). The internet and the investor. *Journal of Economic Perspectives 15*(1), 41–54.

Beshears, J., M. Blakstad, J. J. Choi, C. Firth, J. Gathergood, D. Laibson, R. Notley, J. D. Sheth, W. Sandbrook, and N. Stewart (2024). Does pension automatic enrollment increase debt? evidence from a large-scale natural experiment. Technical report, National Bureau of Economic Research.

Beshears, J., J. J. Choi, D. Laibson, and B. C. Madrian (2013). Simplification and saving. *Journal of economic behavior & organization 95*, 130–145.

Beshears, J., J. J. Choi, D. Laibson, B. C. Madrian, and K. L. Milkman (2015). The effect of providing peer information on retirement savings decisions. *The Journal of finance 70*(3), 1161–1201.

Beshears, J., H. Dai, K. L. Milkman, and S. Benartzi (2021). Using fresh starts to nudge increased retirement savings. *Organizational Behavior and Human Decision Processes 167*, 72–87.

Blumenstock, J., M. Callen, and T. Ghani (2018). Why do defaults affect behavior? experimental evidence from afghanistan. *American Economic Review 108*(10), 2868–2901.

Brody, I., H. Dai, S. Saccardo, K. L. Milkman, A. Duckworth, M. S. Patel, and D. Gromet (2022). Targeting behavioral interventions based on baseline motivation increases vaccine uptake. *Available at SSRN*.

Bryan, C. J., E. Tipton, and D. S. Yeager (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature human behaviour 5*(8), 980–989.

Bursztyn, L., A. L. González, and D. Yanagizawa-Drott (2020). Misperceived social norms: Women working outside the home in saudi arabia. *American economic review 110*(10), 2997–3029.

Bursztyn, L. and D. Y. Yang (2022). Misperceptions about others. *Annual Review of Economics 14*, 425–452.

Caria, S. A. and P. Falco (2022). Skeptical employers: Experimental evidence on biased beliefs constraining firm growth. *Review of Economics and Statistics*, 1–45.

Census (2022). Who has retirement accounts?

Chassang, S., G. Padró i Miquel, and E. Snowberg (2012). Selective trials: A principal-agent approach to randomized controlled experiments. *American Economic Review 102*(4), 1279–1309.

Chen, J. and J. Roth (2023). Logs with zeros? some problems and solutions. *The Quarterly Journal of Economics*, qjad054.

Choi, J. J., E. Haisley, J. Kurkoski, and C. Massey (2017). Small cues change savings choices. *Journal of economic behavior & organization 142*, 378–395.

Choi, J. J., D. Laibson, and B. C. Madrian (2011). $100 bills on the sidewalk : Suboptimal investment in 401(k) plans. Review of Economics and statistics 93(3), 748 − −763.

Choukhmane, T. (2019). Default options and retirement saving dynamics. *Working Pa*.

DellaVigna, S., W. Kim, and E. Linos (2022). Bottlenecks for evidence adoption. Technical report, National Bureau of Economic Research.

DellaVigna, S. and E. Linos (2022). Rcts to scale: Comprehensive evidence from two nudge units. *Econometrica 90*(1), 81–116.

DellaVigna, S. and D. Pope (2018). Predicting experimental results: who knows what? *Journal of Political Economy 126*(6), 2410–2456.

Despard, M., S. Roll, M. Grinstein-Weiss, B. Hardy, and J. Oliphant (2023). Can behavioral nudges and incentives help lower-income households build emergency savings with tax refunds? evidence from field and survey experiments. *Journal of Consumer Affairs 57*(1), 245–263.

Di Tella, R., R. Perez-Truglia, A. Babino, and M. Sigman (2015). Conveniently upset: Avoiding altruism by distorting beliefs about others' altruism. *American Economic Review 105*(11), 3416–3442.

Dizon-Ross, R. (2019). Parents' beliefs about their children's academic ability: Implications for educational investments. *American Economic Review 109*(8), 2728–2765.

Dizon-Ross, R. and A. Zucker (2023). Mechanism design for personalized policy: A field experiment incentivizing exercise. Technical report, Working paper, University of Chicago.

Dur, R., D. Fleming, M. van Garderen, and M. van Lent (2021). A social norm nudge to save more: A field experiment at a retail bank. *Journal of Public Economics 200*, 104443.

Faberman, R. J., A. I. Mueller, A. Şahin, and G. Topa (2022). Job search behavior among the employed and non-employed. *Econometrica 90*(4), 1743–1779.

Fuster, A. and B. Zafar (2023). Survey experiments on economic expectations. In *Handbook of Economic Expectations*, pp. 107–130. Elsevier.

Gabaix, X. (2019). Behavioral inattention. In *Handbook of behavioral economics: Applications and foundations 1*, Volume 2, pp. 261–343. Elsevier.

Haaland, I., C. Roth, and J. Wohlfart (2023). Designing information provision experiments. *Journal of economic literature 61*(1), 3–40.

Hallsworth, M., J. A. List, R. D. Metcalfe, and I. Vlaev (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of public economics 148*, 14–31.

Homonoff, T., R. O'Brien, and A. B. Sussman (2021). Does knowing your fico score change financial behavior? evidence from a field experiment with student loan borrowers. *Review of Economics and Statistics 103*(2), 236–250.

Hubbard, E., P. Matthews, and A. Samek (2016). Using online compound interest tools to improve financial literacy. *The Journal of Economic Education 47*(2), 106–120.

Karlan, D., M. McConnell, S. Mullainathan, and J. Zinman (2016). Getting to the top of mind: How reminders increase saving. *Management science 62*(12), 3393–3411.

Kast, F., S. Meier, and D. Pomeranz (2012). Under-savers anonymous: Evidence on self-help groups and peer pressure as a savings commitment device. Technical report, National Bureau of Economic Research.

Kast, F., S. Meier, and D. Pomeranz (2018). Saving more in groups: Field experimental evidence from chile. *Journal of Development Economics 133*, 275–294.

Kitagawa, T. and A. Tetenov (2018). Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica 86*(2), 591–616.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics 112*(2), 443–478.

Leight, J. and N. Wilson (2020). Framing flexible spending accounts: A large-scale field experiment on communicating the return on medical savings accounts. *Health Economics 29*(2), 195–208.

Levine, P. B. (2022). *A problem of fit: How the complexity of college pricing hurts students—and universities*. University of Chicago Press.

List, J. A. (2022). *The voltage effect: How to make good ideas great and great ideas scale*. Crown Currency.

Loans, D. S. (2023). Annual report.

Loewenstein, G. and D. Prelec (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics 107*(2), 573–597.

Lusardi, A. and O. S. Mitchell (2011). Financial literacy and planning: Implications for retirement wellbeing. Technical report, National Bureau of Economic Research.

Madrian, B. C. and D. F. Shea (2001). The power of suggestion: Inertia in 401 (k) participation and savings behavior. *The Quarterly journal of economics 116*(4), 1149–1187.

OECD, B. I. and P. Policy (2017). Lessons from around the world.

Patterson, J. G. T. H. R. and W. Skimmyhorn (2019). How much to save? decision costs and retirement plan participation.

Review, T. P. (2022). Paying for college.

Roll, S., M. Grinstein-Weiss, E. Gallagher, and C. Cryder (2020). Can pre-commitment increase savings deposits? evidence from a tax-time field experiment. *Journal of Economic Behavior & Organization 180*, 357–380.

Saccardo, S., H. Dai, M. Han, N. Raja, S. Vangala, and D. Croymans (2023). Assessing nudge scalability. *Available at SSRN 3971192*.

Saez, E. (2009). Details matter: The impact of presentation and information on the take-up of financial incentives for retirement saving. *American Economic Journal: Economic Policy 1*(1), 204–228.

Skimmyhorn, W. (2016). Assessing financial education: Evidence from boot camp. *American Economic Journal: Economic Policy 8*(2), 322–343.

Spinnewijn, J. (2015). Unemployed but optimistic: Optimal insurance design with biased beliefs. *Journal of the European Economic Association 13*(1), 130–167.

Stock, J. H. and M. W. Watson (2008). Heteroskedasticity-robust standard errors for fixed effects panel data regression. *Econometrica 76*(1), 155–174.

Tetlock, P. E., B. A. Mellers, and J. P. Scoblic (2017). Bringing probability judgments into policy debates via forecasting tournaments. *Science 355*(6324), 481–483.

Thaler, R. H. and S. Benartzi (2004). Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of political Economy 112*(S1), S164–S187.

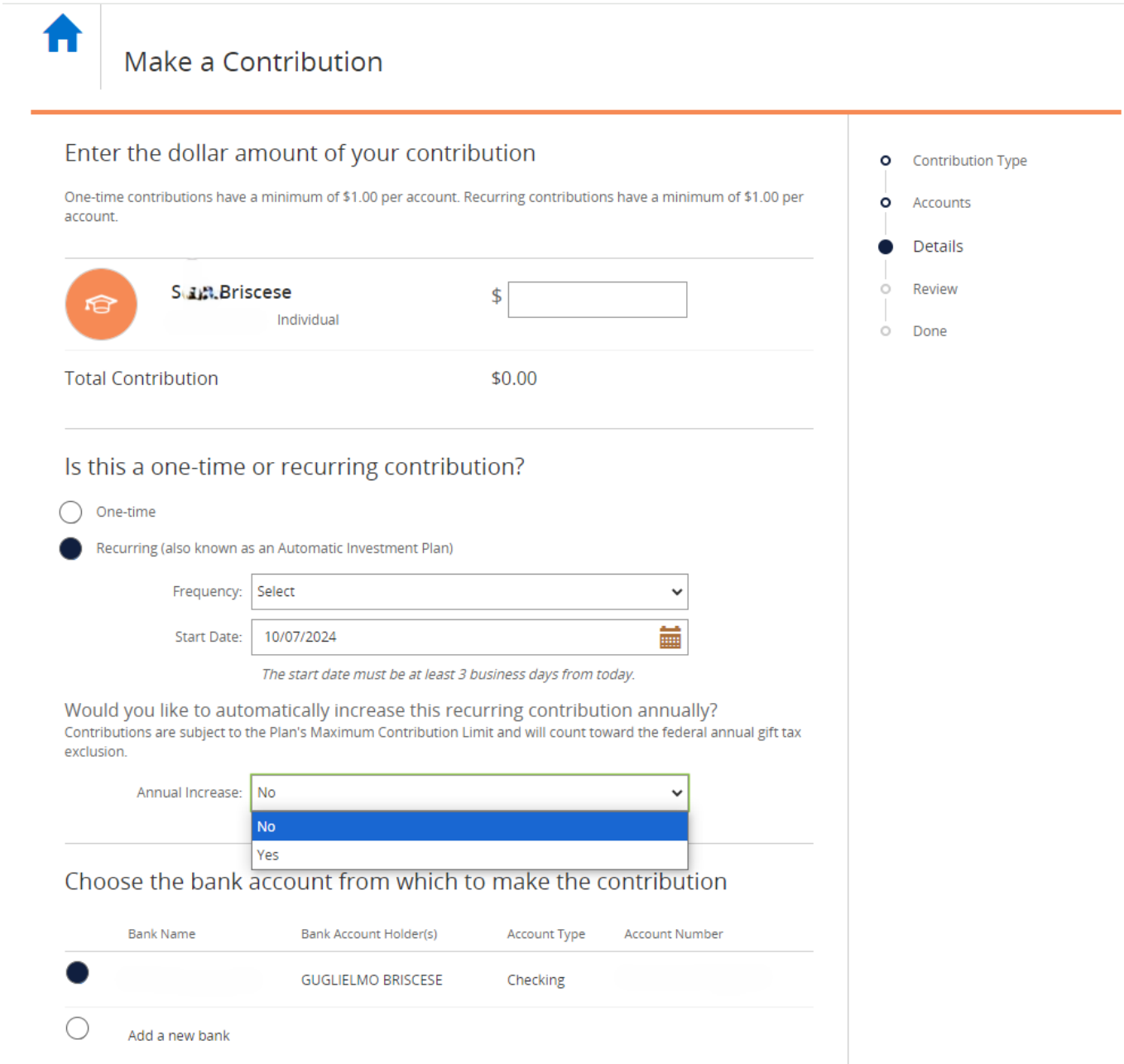
Vissing-Jorgensen, A. (2003). Perspectives on behavioral finance: Does" irrationality" disappear with wealth? evidence from expectations and actions. *NBER macroeconomics annual 18*, 139–194.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

12 Online Appendix

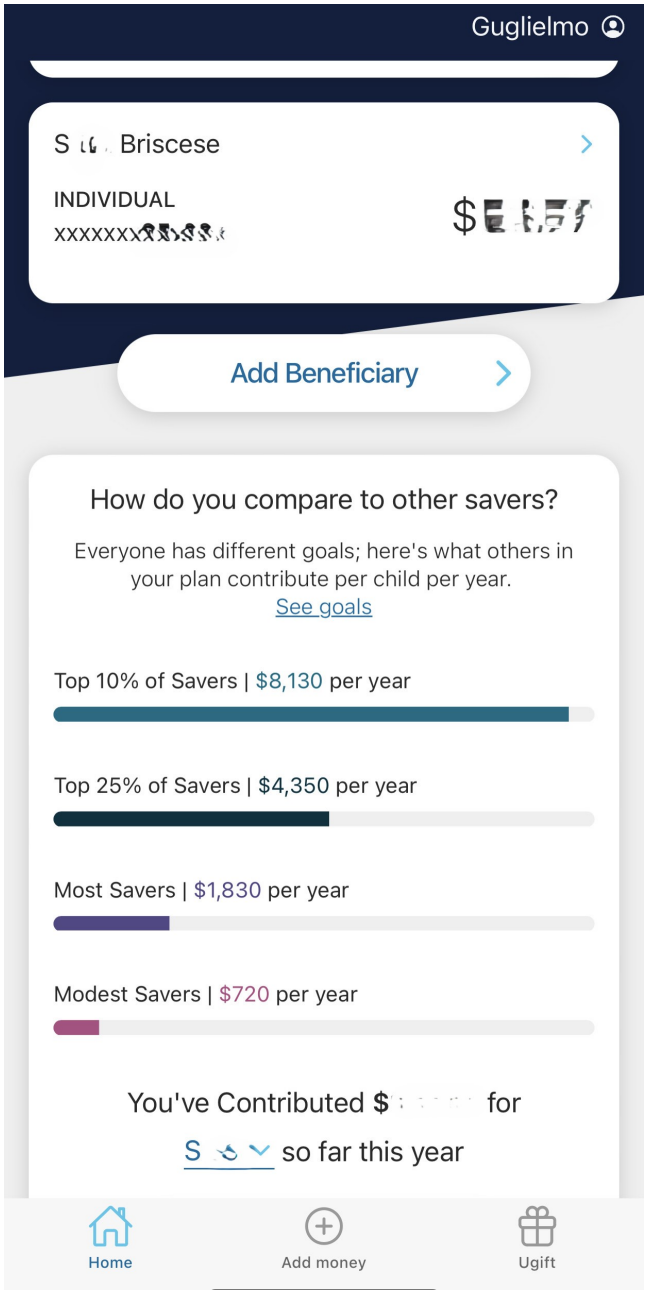
12.1 Policy impact

Figure 18: Post-experiment changes to 529 plans features - Exhibit A



Notes. The figure shows the new feature offered to owners of 529 plans in Illinois that allows for opting into automatically increasing contribution amounts.

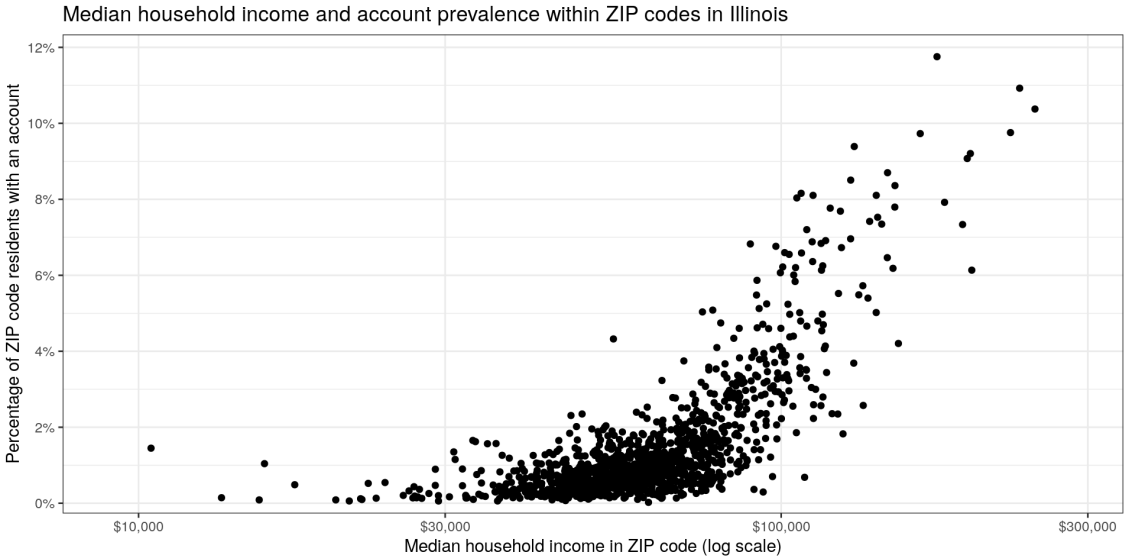
Figure 19: Post-experiment changes to 529 plans features - Exhibit B



Notes. The figure shows the new smartphone app offered to owners of 529 plans in Illinois that allows for personalized messages on the homepage.

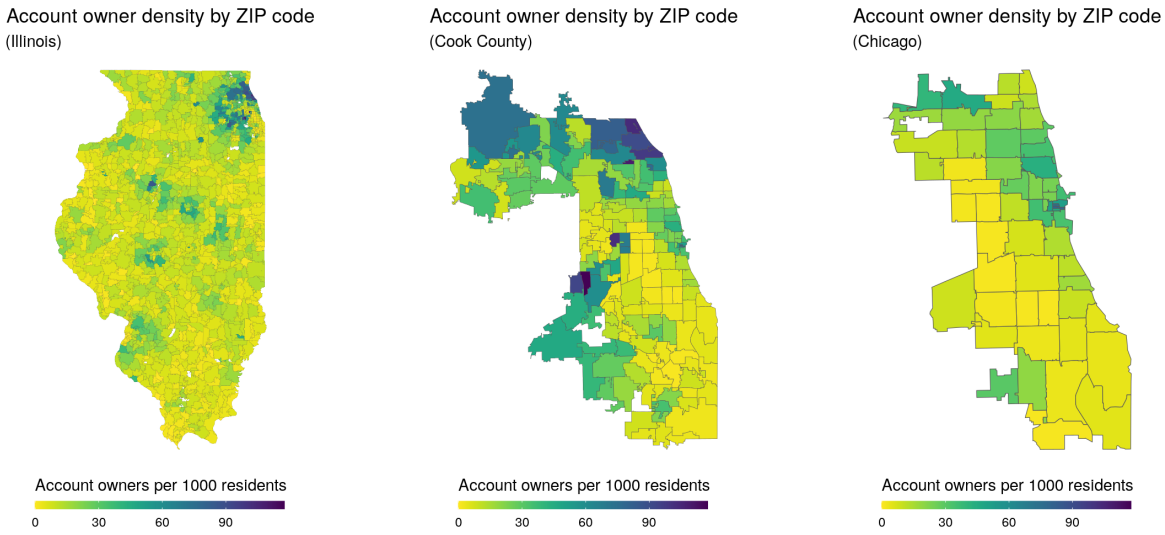
12.2 Descriptive statistics of Illinois 529 plans

Figure 20: IL 529 take up rate by income



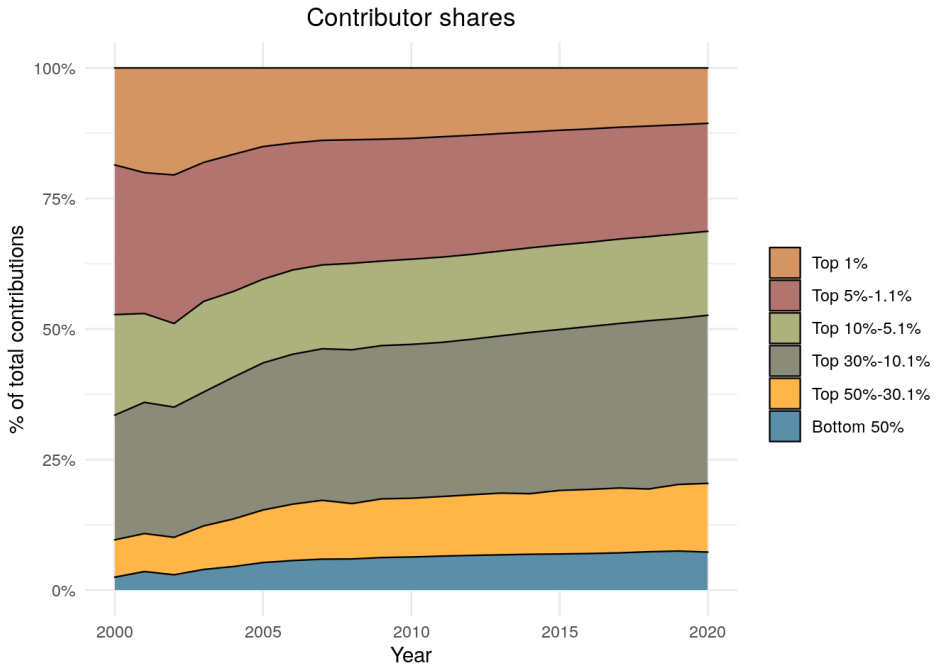
Notes. The graph shows the ZIP code's household median income (on the x-axis) and the percentage of ZIP code residents with an account (on the y-axis) using 2000-2020 529 data and 2020 Census data.

Figure 21: Geographical distribution of IL 529 plans ownership



Notes. The maps use population estimates from the 2019 ACS data. There are 263,122 account owners represented in the Illinois map. Density is defined as: [(number of account owners in ZIP code / number of residents in ZIP code) * 1000.]

Figure 22: Saving distribution in IL 529 plans



Notes. The graph shows the share of total assets under management in the Illinois 529 program in any given year owned by account owners.

12.3 Study materials

A copy of the survey invite email sent by the partner agency to all account owners can be found here.

A copy of entire survey can be found here.

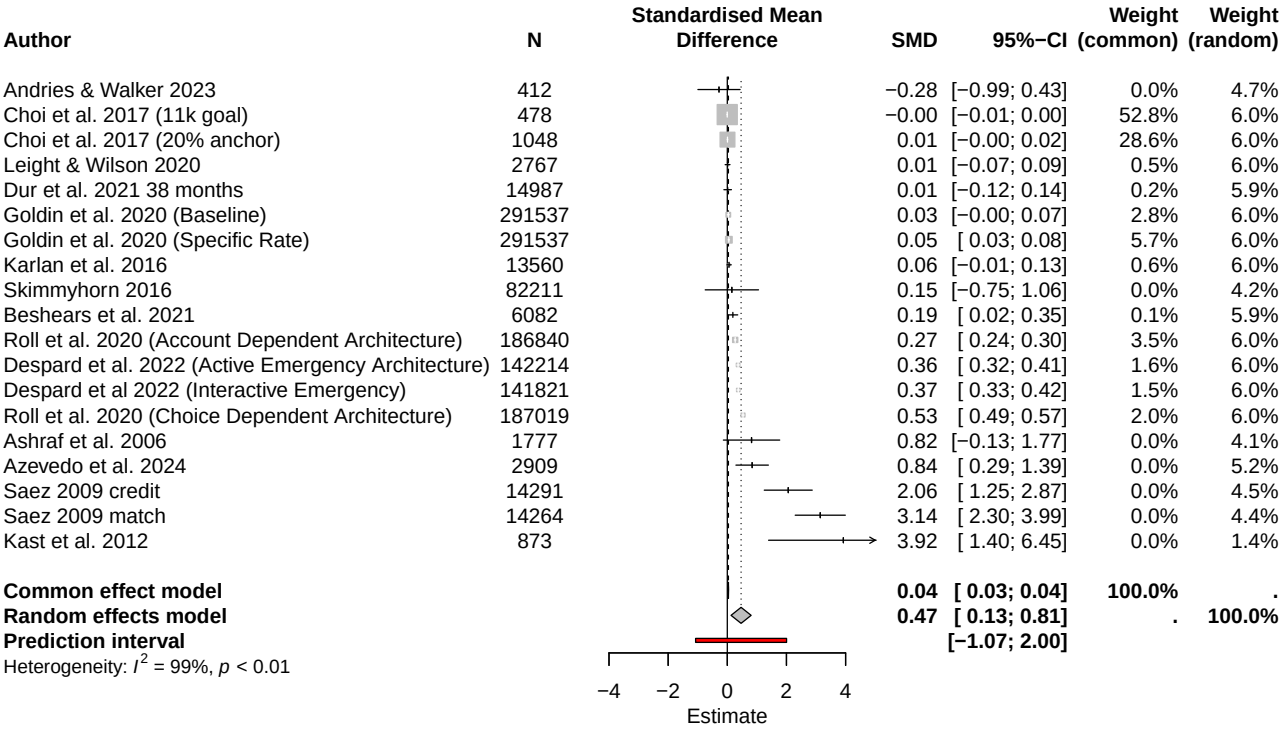
12.4 Meta-analysis - For online publication only

In this Appendix, we provide a summary of the meta-analysis used to identify patterns in the efficacy of nudge interventions to increase savings. As with most meta-analysis, this section of the manuscript is intended to provide the basis for a regularly update online repository that can inform researchers and policymakers on the reliability of behavioral change interventions in this policy context. We restricted our survey of the literature to randomized controlled trials (RCTs) that were published after 2004, using Thaler and Benartzi (2004)'s publication date as a cut-off point. We prioritized interventions that were comparable to the one we aimed to implement with our partner agency - that is, nudge interventions that consisted in light-touch scalable behaviorally-informed communications that encouraged greater savings. While some studies focus on encouraging greater savings among individuals without a savings account, we solely examine the treatment effect on savings rates (as opposed to take-up rates).

Studies selection. We used searches of Google Scholar, ScienceDirect, and EBSCO using the following keywords: "nudges" or "intervention" AND "savings" AND "accounts" AND "randomized" OR "field experiment." Our search yielded 20 papers with a total sample of 862,945. Fromt this set of studies, we excluded 6 that failed to report needed information to conduct the meta-analysis, namely they were missing information on standard errors, average savings in the control group, or the sample size. This left us us with 14 studies and 19 unique treatments, with a total sample size of 761,481. These studies range in publication from 2006 to 2023 and come primarily from a economics journals. These studies are: Azevedo et al. (2024); Despard et al. (2023); Andrieş and Walker (2023); Beshears et al. (2021); Dur et al. (2021); Roll et al. (2020); Leight and Wilson (2020); Patterson and Skimmyhorn (2019); Choi et al. (2017); Skimmyhorn (2016); Karlan et al. (2016); Kast et al. (2012); Saez (2009); Ashraf et al. (2006).

Effect sizes. Like other meta-analyses, we standardized the effect size across studies using the study's pooled standard error (Cohen 1988) and used a random effects model. We found that nudge interventions have a medium-sized positive effect on individual savings rates by 0.47 standard deviations compared to a control group (d= 0.47, CI = [0.13,0.81], I^2= 99, p <0.01) as shown in Figure 23). We also find high heterogeneity across studies.

Figure 23: Meta-analysis forest plot



Notes. The figure shows the standardized mean difference of each treatment arm (in parenthesis in the first column) of each selected study (authors listed in the first column) and the results of the random effects model. The horizontal lines represent the 95% confidence interval, with an arrow flagging when the interval extending beyond the limits of the figure.

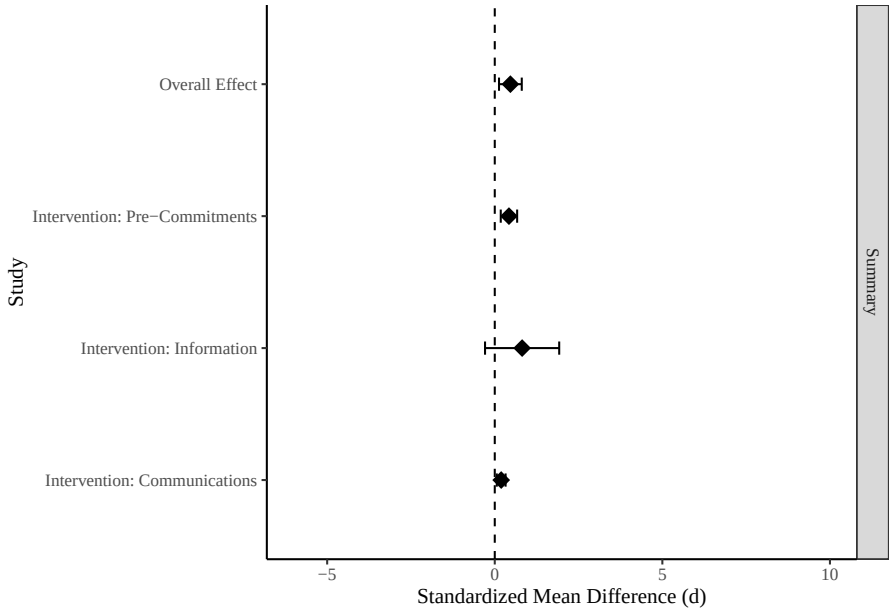
To improve comparability of the selected studies, we group them by the nature of the behaviorally-informed communication being used into four categories: communication (studies that used behaviorally-informed nudges), information (studies that offered a more informative explanation in the form of financial literacy or advice), commitment (studies that suggested an anchoring pre-specified savings amount), and social comparison (studies that use norms in their messaging). Informational nudges (d = 0.82, I^2= 99%) have a larger effect than communication (d= 0.19, I^2= 97%) and commitment (d = 0.42, I^2= 98%). The persistence of high heterogeneity can arguably be attributed to the different factors influencing nudge interventions such as tested population, timing, and delivery method (see Figure 24 for a visual summary of the grouped effects).

Table 22: Selected studies for meta-analysis

Citation	Category	N	Treat. eff. compared to contr.	S.E.
Leight & Wilson 2020	communications	2767	0.01	0.04
Azevedo et al. 2024	communications	2909	0.84	0.28
Karlan et al. 2016	communications	13560	0.06	0.04
Despard et al. 2022 (Active Emergency)	communications	142214	0.36	0.02
Despard et al 2022 (Interactive Emergency)	communications	141821	0.37	0.02
Goldin et al. 2020 (Baseline)	communications	291537	0.03	0.02
Goldin et al. 2020 (Specific Rate)	communications	291537	0.05	0.01
Beshears et al. 2021	communications	6082	0.19	0.08
Kast et al.2012	communications	873	3.92	1.29
Saez 2009 match	information	14264	3.14	0.43
Saez 2009 credit	information	14291	2.06	0.41
Andreis & Walker 2023	information	412	-0.28	0.36
Choi et al. 2017 (20% anchor)	information	1048	0.01	0.01
Choi et al. 2017 (11k goal)	information	478	0.00	0.00
Skimmyhorn 2016	information	82211	0.15	0.46
Roll et al. 2020 (Choice Depen. Architect.)	pre-commitments	187019	0.53	0.02
Roll et al. 2020 (Account Depen. Architect.)	pre-commitments	186840	0.27	0.02
Ashraf et al. 2006	pre-commitments	1777	0.82	0.49
Dur et al. 2021 38 months	social-comparison	14987	0.01	0.07

Notes. The table reports the list of selected studies used in the meta-analysis, with the authors and year of publication listed in the first column, together with the treatment arm selected for comparability where needed.

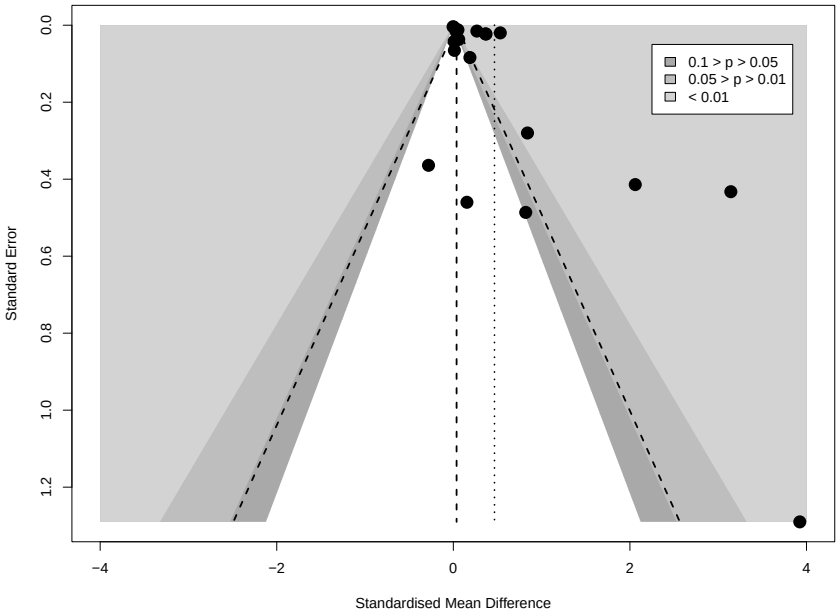
Figure 24: Meta-analysis forest plot



Notes. The figure shows the standardized mean difference of nudge interventions grouped by type, with their related 95% confidence interval.

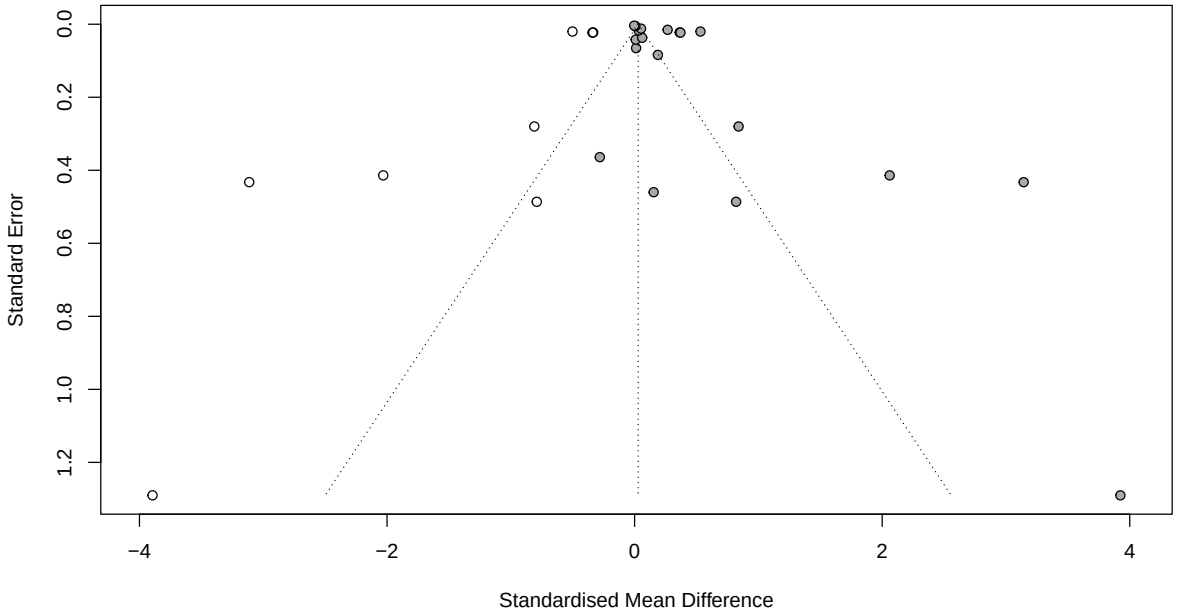
Publication Bias. In this last section, we test for possible publication bias by visual means of a funnel plot. Should the literature be free from this issue, we would observe an even distribution amongst the unshaded funnel region which is demarcated by the dotted line representing the 95% confidence interval. From Figures 25 and 26 we see that, in line with the results by (DellaVigna and Linos, 2022), also this subset of selected studies isn't immune from publication bias.

Figure 25: Meta-analysis funnel plot



Notes. The figure shows the standardized mean difference of nudge interventions grouped by type, with their related 95% confidence interval.

Figure 26: Meta-analysis trim filled funnel plot



Notes. The figure shows the standardized mean difference of nudge interventions grouped by type, with their related 95% confidence interval.

The funnel plots also indicate significant asymmetry with positive results. While it is possible that the specific nature of the outcome variable (i.e., savings) reduces the chances of seeing negative effects (which may be mostly zero when null), it may also be representative of publication bias. Figure 26 identifies the need for eight studies to achieve symmetry of probable effects. Adjusted for these eight studies, the effect size falls to 0.03 standard deviations (95% CI = [-0.4186, 0.4760], I^2= 99%). When adjusting for publication bias, we find a much smaller, statistically insignificant effect size.