

How Effective Is (More) Money? Randomizing Unconditional Cash Transfer Amounts in the US*

Ania Jaroszewicz, Oliver P. Hauser, Jon M. Jachimowicz, and Julian Jamison

July 20, 2024
[Click here for the most recent version](#)

Abstract

We randomized 5,243 US individuals in poverty to receive a one-time unconditional cash transfer (UCT) of \$2,000 (two months' worth of total household income for the median participant), \$500 (half a month's income), or nothing. We measured the effects of the UCTs on participants' financial well-being, psychological well-being, cognitive capacity, and physical health through surveys administered one week, six weeks, and 15 weeks later. While bank data show that both UCTs increased expenditures, we find no evidence that (more) cash had positive impacts on our pre-specified survey outcomes, in contrast to the incentivized predictions of both experts and a nationally representative sample of laypeople. We test several explanations for these unexpected results, including via two sub-experiments embedded in our trial. The data are most consistent with the notion that receiving some but not enough money made participants' (unmet) needs more salient, which caused distress. We develop a model to illustrate how receiving cash can sometimes also highlight its absence. (JEL: C93, D91, G50, I30)

Living in poverty has been linked to a wide range of detrimental outcomes, including worse psychological well-being (Haushofer and Fehr, 2014; Ridley et al., 2020), poorer physical health (Braveman et al., 2010), and more limited cognitive capacity (Mani et al., 2013; Schilbach, Schofield and Mullainathan, 2016). Whether explicitly or implicitly, many researchers and policymakers have argued that providing people with more money—e.g., in the form of unconditional cash transfers (UCTs), “no strings attached” financial payments—should help address these issues and generally improve the recipients’ outcomes. Indeed, there is good reason to think that this should be the case: Prior research in low-income countries has shown that cash transfers often (though not always) improve individuals’ outcomes, for instance increasing consumption and food security.¹

*Jaroszewicz (corresponding author): University of California San Diego, aniaj@ucsd.edu. Hauser: University of Exeter, o.hauser@exeter.ac.uk. Jachimowicz: Harvard Business School, jjachimowicz@hbs.edu. Jamison: University of Oxford & University of Exeter, J.Jamison@exeter.ac.uk. This paper was previously circulated under the title “Cash Can Make Its Absence Felt: Randomizing Unconditional Cash Transfer Amounts in the US.” This RCT was registered as AEARCTR-0006149 (Jaroszewicz, 2020) and obtained IRB approval (Harvard IRB20-0813 and IRB20-1473). First and foremost, we would like to thank our field partners, without whom this trial would not have been possible. We would also like to thank Christine Exley; Johannes Haushofer; Maurizio Montone; Sally Sadoff; Catherine Thomas; Keela Thomson; attendees of the AFE, AOM, Boulder Consumer Financial Decision-Making, European and North American ESA, IPA-GPRL, PEP, SJDM, and SPSP conferences; seminar and lab participants at Exeter, Harvard, Johns Hopkins, the LSE, NUS, UCSD, and Utrecht; and especially Emily Hauser and Joel Levin for their invaluable feedback. Matthew Freedman, Matthew Higgins, Michaela Moulaison, Adriàan Rodriguez, Sandhya Srinivas, Bonnie Tacheron, Kevin Wong, and in particular Arvo Muñoz Morán provided excellent research assistance. This study was funded by our field partner.

¹For largely positive effects, see e.g. Baird, McIntosh and Özler (2011); Miller, Tsoka and Reichert (2011); Robertson et al. (2013); Blattman, Fiala and Martinez (2014); Haushofer and Shapiro (2016); Hidrobo, Peterman

However, there is also reason to think that UCTs may not have uniformly positive effects, especially if they are one-off and of the relatively modest amounts often provided in high-income countries.² First, lacking financial resources can produce complicated and persistent issues, such as isolation and limited access to opportunities, which may not readily be addressed through a one-time UCT. Second, increases in objective wealth may not necessarily correspond to increases in subjective wealth and well-being, both because the correlation between these is only moderate (Gasiorowska, 2014), and because opportunities for upward mobility may at times generate their own social and psychological challenges (Sorokin, 1959; Friedman, 2016; Präg, Fritsch and Richards, 2022).³ This opens the possibility for somewhat more nuanced effects. For instance, perhaps UCTs cannot address deep problems (e.g., depression), but are suitable for simpler ones (e.g., paying for groceries); or perhaps they do not have measurable short term benefits, but help people avoid costly debts in the long run.

To better understand the effects of UCTs on low-income individuals, we collaborated with a non-profit organization to run a preregistered longitudinal field experiment in the US.⁴ This study, conducted between July 2020 and May 2021, randomized 5,243 low-income Americans to receive either (1) \$0 (hereafter: “Control”; $N = 3,170$), (2) a one-time UCT of \$500 (roughly half a month’s worth of median total household income; $N = 1,374$), or (3) a one-time UCT of \$2,000 (two months’ worth of income; $N = 699$). All participants took a baseline survey before being randomized and were invited to take post-treatment surveys one week, six weeks, and 15 weeks after the cash transfer. These surveys measured the participants’ financial well-being (e.g., subjective financial stress, liquidity constraints), psychological well-being (e.g., depression, agency), cognitive capacity (e.g., a fluid intelligence measure, the extent to which the participant thought about money), and physical health (e.g., food security, sleep quality). We summarize these four outcome categories as four separate indices, which together serve as our primary outcome variables. We also observe bank account balances and financial transactions during the trial period for the

and Heise (2016); Blattman, Jamison and Sheridan (2017); Handa et al. (2018); Baird, McIntosh and Özler (2019); Christian, Hensel and Roth (2019); Haushofer et al. (2019); Brooks et al. (2022); Karlan et al. (2022); Londoño-Vélez and Querubín (2022); Banerjee et al. (2023); Cañedo, Fabregas and Gupta (2023); Haushofer, Mudida and Shapiro (2023); Richterman et al. (2023); Wollburg et al. (2023); Aggarwal et al. (2024) and Gupta et al. (2024). Some mixed and null outcomes have also been reported—see, e.g., Berge, Bjorvatn and Tungodden (2015); Field and Maffioli (2021); Andersen, Kotsadam and Somville (2022); Banerjee et al. (2022); Bartos et al. (2022); Hussam et al. (2022) and Aiken et al. (2023).

²For instance, in 2020-2021, the US federal government gave most Americans UCTs in the form of three “economic stimulus checks,” totaling \$1,200, \$600, and \$1,400, respectively. While substantive for such a widespread policy, these amounts are just a fraction of what is often given in lower income countries in terms of purchasing power. According to calculations by Dwyer, Stewart and Zhao (2023), the average cash transfer amount in the Global South, \$1,414, corresponded to 37% of the recipient’s country’s median per capita annual income in 2021 dollars. That same percentage in the US would amount to almost \$14,000—likely a prohibitively large sum for many programs. (Economic stimulus check amounts calculated from <https://www.census.gov/quickfacts/fact/table/US/SEX255219>; accessed 6 October 2023).

³This work, largely stemming from sociology, has argued that such opportunities may generate tensions about which community one belongs to (Lee and Kramer, 2013; Curl, Lareau and Wu, 2018), uncertainty about one’s identity (Hurst, 2010; Destin and Debrosse, 2017), and guilt in asking why the self was afforded opportunities that others were not (Covarrubias, Romero and Trivelli, 2015).

⁴Preregistration and preanalysis plan: <https://www.socialscienceregistry.org/trials/6149>

43% of participants who opted into providing that data. These data allow us to measure when and how the cash transfers were spent, as well as how much money was saved and for how long.

To measure people’s priors on the potential effects of these cash transfers, we conducted an incentivized prediction study (Dreber et al., 2015; DellaVigna, Pope and Vivaldi, 2019) concurrently with data collection for the main study.⁵ We recruited two samples, one of social scientists and policymakers (“experts”; $N = 477$) and another that was representative of the US population on standard demographics (“laypeople”; $N = 971$). Both groups made incentivized predictions about the outcomes of the field experiment, estimating the standardized effect sizes of both treatments (relative to Control) on each of the four survey outcome indices at each of the post-treatment surveyed time points. We find that both experts and laypeople predicted positive effects of both cash amounts on each of the indices at each time point, believing that the \$500 group would outperform the Control group, and that the \$2,000 group would outperform the \$500 group. Average effect size predictions (relative to Control) ranged from 0.16 to 0.65 SDs , depending on the treatment, index, and time point. For instance, experts predicted an effect size of 0.49 SDs for the \$2,000 group on the financial index one week after cash receipt.

In reality, however, our field experiment results reveal no significant positive effects of either cash amount on any of the preregistered survey outcomes. This is despite the fact that bank data show that participants spent their money well within the survey time periods, and seemingly primarily on bills and other necessities. In fact, at every post-treatment survey time point, both cash groups reported significantly *worse* outcomes than the Control group on the financial, psychological, and health survey indices (and no significant differences on the cognitive capacity index). We also find no differences between the \$500 and \$2,000 groups for any of the indices at any time point, and generally find few differences across the post-treatment time periods. These non-positive effects of either cash amount are robust to a wide range of alternative specifications. Subsequent analyses reveal that what appears non-positive on average may be the result of two different forces among our composite indices. For some items, there are negative effects of cash—these are concentrated among self-reports of more *subjective* experiences of outcomes (e.g., how the participant evaluated a certain element of their lives). For other items, there are no detectable effects of cash—these are concentrated among self-reports of more *objective* outcomes (e.g., dollar amounts or the number of days in the past week that an event occurred). This pattern suggests that the cash did not actually produce worse outcomes in some objective sense, but nevertheless made some recipients *feel* worse.

What can explain the lack of positive effects of cash? First, we examine attrition carefully. We observe relatively high response rates of post-treatment survey data for a UCT trial in a high-income country, with some variation across conditions: 80% of Control group participants, 90% of the \$500 group participants, and 88% of the \$2,000 group participants took at least one post-treatment survey. One could argue that, if the unobserved Control group participants had particularly bad post-treatment outcomes and/or the unobserved cash group participants had par-

⁵At the point of launching the prediction study, we had received only 1% of the post-treatment surveys from the main study. Appendix Section A provides more details on the methods and results of the prediction study.

ticularly good outcomes, our estimates could be biased downwards, masking a potentially more positive effect of cash. (Arguably, the opposite could be true instead: if the unobserved Control group participants had particularly good post-treatment outcomes and/or unobserved cash group participants had particularly bad outcomes, our estimates could be biased upwards, and the true effect of cash would be less positive than it already is.) To test whether and how attrition could have affected the direction and magnitude of our effects, we conduct a wide range of analyses, imputation exercises, and bounding exercises. We find that, in specifications where we observe negative effects of receiving cash, attrition could have made them appear more negative than they might otherwise have been. However, these analyses also demonstrate that it is *highly unlikely* that the effects of cash could have been meaningfully positive. In other words, even once we take into account attrition, we are confident that the effects of receiving (more) cash are not positive on the outcomes we study.

We then proceed to investigate seven potential mechanisms to explain why we may not have seen positive effects, including through two sub-experiments we embedded in the surveys ($N = 2,423$ and $N = 2,474$). We test for strategic distortion of responses, reference dependence, harmful spending, mismatched expectations, negative inferences about the self, and declining social relationships. Although some of the tests provide possible evidence consistent with some of these mechanisms, we rule out most of them with reasonable confidence.

Instead, our data is most consistent with the following mechanism. Receiving cash may have made participants consider the ways in which they could spend that cash—i.e., think more deeply about existing financial obligations and potentially uncover new ones. This, in turn, could have caused distress, particularly if they discovered that these obligations were larger than previously thought and the windfall is insufficient to address them. In support of this mechanism, we find that participants who received cash thought more about money and how to spend it, reported needing more money to meet their household’s obligations across a wider range of spending categories, and felt more overwhelmed by the needs of people outside their household. Depending on the analyses, these variables either partially or fully statistically mediate the effect of cash on the survey outcomes. This mechanism is consistent with literature documenting information aversion, and in particular that focusing attention on one’s bad financial state can be unpleasant (Karlsson, Loewenstein and Seppi, 2009; Andries and Haddad, 2020).

We rationalize these findings through an economic model that takes as a starting point an agent who chooses to optimally pay down obligations (e.g., debt) over two periods. Agents choose to be passive or active in managing their obligations, where passively managing these obligations avoids the psychological and economic costs of active management but prevents agents from noticing any financial shocks. We model the agent’s best response to an exogenous windfall—akin to our cash treatments—and find that they are more likely to choose to actively manage more of their obligations when receiving the windfall. This, however, leads agents to experience lower utility in the first period because, for a non-trivial range of model parameters, the obligations they uncover are larger than they expect and the windfall is insufficient to address all obligations. Our model

further hypothesizes that while agents initially feel worse, they experience higher lifetime utility from having reduced obligations earlier. The model also predicts that larger windfalls may attenuate or even reverse negative utility experienced in the first period.

Our results allow us to make several contributions to the literature. The first relates to providing more rigorous evidence on the effectiveness of UCTs in high-income countries, particularly during challenging economic conditions.⁶ In high-income countries, existing studies have often been limited by non-experimental methods, relatively small samples, and/or outcome measures that are restricted in frequency or scope.⁷ Moreover, although most research in high-income countries has suggested positive effects of cash on various outcomes, particularly for children, other studies have documented mixed or no detectable effects.⁸ These methodological constraints and mixed results have left open various questions about which outcomes UCTs may affect, when, and why (or why not). Indeed, in the two studies most similar (and concurrent) to ours (Jacob et al., 2022; Pilkauskas et al., 2023), low-income US households received \$1,000 UCTs during the pandemic and reported material hardship and mental health (and, in the case of Pilkauskas et al. (2023), parenting, child behavior, and partner relationships) between one and three months later. Neither study finds any average treatment effects of cash, though Pilkauskas et al. (2023) find evidence of reduced material hardship among the poorest participants in the sample. One possibility is that somewhat more money would have produced more positive effects. Another is that measuring outcome variables on a shorter time frame (less than one month) would have uncovered positive results. A third is that there might have been positive effects on other outcomes they did not test. Our experiment—which randomized UCT amounts⁹ and measured a particularly wide range of outcomes over time using both surveys and administrative data—allows us to more thoroughly and rigorously answer these open questions. We address several of these alternative explanations at once: even doubling the size of UCT payments (relative to past work) and measuring outcome variables only one week later shows no positive effects, and in fact effects may reasonably be negative or indistinguishable from

⁶Most empirical evidence primarily stems from lower-income countries (see Footnote 1). One reason why we might expect UCTs to have more positive impacts in lower- (vs. higher-) income countries is that the needs are greater on average and researchers' money goes further (Dwyer, Stewart and Zhao, 2023). Though direct comparisons are difficult, related possibilities are that recipients in low-income countries may be more likely to spend the money on durables (e.g., a new roof) rather than non-durables (e.g., rent, groceries), and they may be more likely to spend the money in entrepreneurial ways.

⁷Non-experimental studies include those examining the effects of receiving government benefits or payments (e.g., Akee et al. (2010); Milligan and Stabile (2011); Dahl and Lochner (2012); Watson, Guettabi and Reimer (2019); Erten, Keskin and Prina (2022); Kovski et al. (2023); Pignatti and Parolin (2023); and Silver and Zhang (2023)) and the effects of winning a lottery (e.g., Kuhn et al. (2011); Cesarini et al. (2016); Kent and Martínez-Marquina (2022) and Golosov et al. (2024)). Experimental studies include Salkind and Haskins (1982); Persaud et al. (2021); Dwyer and Dunn (2022); Troller-Renfree et al. (2022); Liebman et al. (2022); Dwyer et al. (2023), and Kluender et al. (2024).

⁸Those identifying mixed or no effects include Gardner and Oswald (2007); Evans and Moore (2011); Carvalho, Meier and Wang (2016); Price and Song (2018); Persaud et al. (2021); Jacob et al. (2022); Pilkauskas et al. (2023); Dwyer et al. (2023); Silver and Zhang (2023); Aizer et al. (2024); Kluender et al. (2024); and Gennetian et al. (2024). There are a number of differences between the high-income country studies finding positive and mixed/no effects that could explain the variability in conclusions, including the amount of money received, the outcomes measured, the payment frequency, the subject populations, and the recipients' expectations about the funds.

⁹Almost all prior work evaluates a single amount of a cash transfer versus no cash transfer. One exception is the negative income tax experiments conducted between 1968 and 1982 in the US and Canada (Widerquist, 2005).

zero across a wide range of outcomes. Overall, these findings suggest a narrower range of possible circumstances under which one-off UCTs could have detectably positive effects in similar contexts. Moreover, when viewing these results in conjunction with the prediction study, in which even experts were quite optimistic about the effectiveness of cash, our data challenge what may have been widespread and overly optimistic priors about the effects of UCTs in high-income countries. Taken together, the results suggest that our posteriors about the effectiveness of similar cash transfer policies in similar settings should be somewhat tempered.

The second contribution relates to the psychology of poverty and scarcity, the feeling that one has fewer resources than one needs (Mullainathan and Shafir, 2013). As previously mentioned, prior research has argued that having insufficient money can impose a range of emotional and cognitive burdens (Shah, Mullainathan and Shafir, 2012; Schilbach, Schofield and Mullainathan, 2016; Shah, Mullainathan and Shafir, 2019; Ridley et al., 2020; Jachimowicz et al., 2021; Kaur et al., 2021). Contrary to the predictions of this work, we find that providing additional resources does not necessarily alleviate these adverse effects and may in fact actually produce additional psychological strain for some.¹⁰ Moreover, our work offers one potential explanation for when and why this may be the case: we argue that receiving (insufficient) money may in some cases bring to mind not only the needs and obligations that it *can* address, but also those that it *cannot*. Our results suggest that people’s baseline perceptions of their obligations may at times capture only a subset of their actual obligations, and receiving a one-off cash transfer may prompt them to engage with their financial situation more deeply and uncover more obligations. Viewed through the lens of our model, one way to interpret the experimental results is that, even though people were seemingly able to use the cash to address their needs, the psychological and transactional costs of uncovering a fuller but potentially unexpectedly bad view of their finances also caused some psychic disutility. Indeed, our findings and model are consistent with a contemporaneous paper testing the effects of medical debt relief, and awareness of that relief, on a range of outcomes (Kluender et al., 2024). The authors find no average effect of debt relief on mental health and even observe detrimental effects for those who were randomly assigned to receive phone calls drawing attention to the treatment, which (as noted by the authors) is consistent with our proposed mechanism. Our findings contribute to the literature on the psychology of poverty by beginning to uncover how and why even well-intentioned poverty-relief policies (e.g., UCTs to low-income households or debt relief) may in some cases generate unintended negative effects on well-being.

We describe the experiment methods in Section 2, then provide an overview of the administrative data (Section 3) and survey results (Section 4). In Section 5, we discuss possible explanations for the unexpected effects and present a model to help illuminate the explanation for which we have the most evidence. We conclude in Section 6 with a discussion of learnings and policy implications.

¹⁰Other work finding little evidence for some of the hypothesized negative effects of scarcity include Carvalho, Meier and Wang (2016); Camerer et al. (2018); O’Donnell et al. (2021); de Bruijn and Antonides (2022) and Szasz et al. (2023).

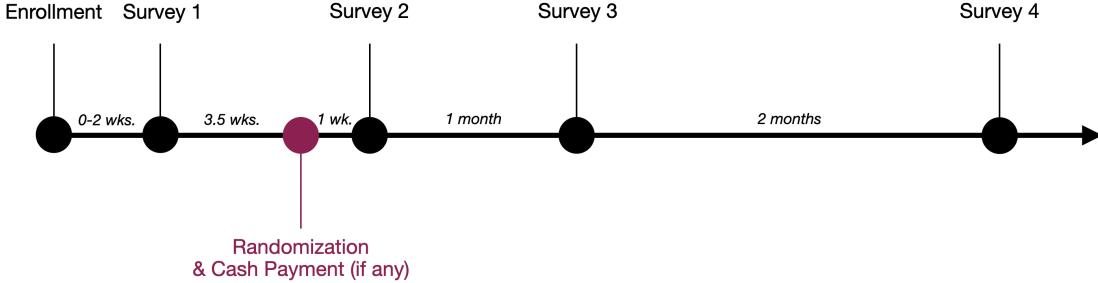


Figure 1: Timeline for the field experiment.

2 Methods

The study was conducted in the US from July 2020 to May 2021, soon after the start of the COVID-19 pandemic. It was run in close collaboration with a national non-profit organization that specializes in providing low-income people cash transfers. Figure 1 shows a timeline of the study.

2.1 Enrollment

All participants had previously applied to the organization for COVID-19 relief funds. About 43% of participants received funds from the non-profit before the trial began. Of that 43%, 99% had received \$500, with a median of 243 days between their pre-trial payment and their trial start date.¹¹

The non-profit recruited participants from among this pool of applicants by email. The email advertised up to \$100 in survey participation payments (\$20 for completing each of four surveys, plus an extra \$20 if they completed all four). It also advertised a chance of winning \$1,000 after the study was over (and thus, after all outcome measures were collected). The recruitment email did not mention the possibility of receiving \$500 or \$2,000, although some participants seemed to suspect or hope that they might receive money; see Appendix Section E.1. Enrollment was conducted across seven different “waves,” such that participants in each wave began the study on different dates, but the treatment group randomization ratios and the time gaps between intervention points did not change. All participants provided informed consent.

Within a few weeks of enrolling, participants were sent the first of four surveys: the baseline or “t1” survey (see Section 2.3 for a description of the surveys). Only participants who completed this survey were subsequently randomized. Column “t1” of Appendix Table H.1 shows the enrollment by wave.

2.2 Randomization and treatment

Roughly 3.5 weeks after completing the baseline survey, participants were randomized into one of three treatment arms. The Control group ($N = 3,170$) did not receive any additional funds beyond

¹¹Less than 1% of participants received a payment in the four months before they started the trial. Excluding them does not qualitatively change the results.

the survey completion payments (and potentially payments for providing their bank data; see Section 2.4). The \$500 group ($N = 1,374$) received \$500 as a one-time UCT—the equivalent of $1.2 \times$ monthly median earned household income and about half of monthly median total household income (i.e., earned plus unearned income). The \$2,000 group ($N = 699$) received \$2,000: approximately $4.8 \times$ monthly median earned household income and $2 \times$ monthly median total household income.¹² Recipients could use the money however they wanted. The unequal sizes of the treatment groups were based on a calculation of how to maximize power given unequal treatment group costs and a fixed budget (List, Sadoff and Wagner, 2011). In addition, at the non-profit’s request, participants were cross-randomized to either receive or not receive access to certain new design features of the non-profit’s online platform, which were separate from the financial aspects of the platform. As specified in our preanalysis plan, we do not analyze the effects of access to these additional features, but control for it in our regressions.¹³

When a UCT or survey payment was sent, participants were notified of it by email. The UCT email additionally informed participants that they had been randomly chosen to receive the money. All payments, including UCTs and survey payments, were sent in one of two ways. For 90% of payments, the non-profit placed the money onto its online platform and allowed the participant to “pull” that money into their own external account (e.g., checking account). Initiating such a pull was a simple process, requiring just a few clicks, an indication of how much money they wanted to pull, and (optionally) what they intended to use the money for. Pull requests were typically fulfilled within a couple of days. For the remaining 10% of payments, the money was “pushed” directly to participants’ external accounts. Participants could choose whether they wanted to receive it in a bank account, on a virtual payment card, or on a physical payment card.¹⁴

2.3 Surveys

We administered four surveys. As mentioned above, the “t1” or “baseline” survey occurred 3.5 weeks before the cash transfers. The next three (the “post-treatment surveys”) occurred roughly one week (“t2”), one month and one week (“t3”), and three months and one week (“t4”) after the cash transfer. This schedule meant that participants were always taking the surveys at the same time of the month, which we believed would minimize noise in survey responses by ensuring that

¹²These funds were non-taxable. However, if participants reported the payments, their welfare benefits could be affected. Participants were encouraged to speak with their case managers or benefits counselors to understand how these payments could affect welfare benefit receipt.

¹³In any case, access to these additional features had no effect on any of our outcomes.

¹⁴In some cases, the non-profit sent payments but the participant did not receive it in their external account. This could happen for one of two reasons. First, money on the online platform may have not been pulled into the external account, which could reflect the participant’s intentional decision (e.g., holding a “rainy day fund”) or an unawareness of the money. Second, the money might have been pushed or attempted to have been pulled, but failed to arrive in the bank account due to errors such as the participant inputting their bank account information incorrectly or the account closing. Although our primary analysis uses an intent-to-treat—and thus we analyze cash group participants as such regardless of whether they were sent, received, or used the money—we can verify that the non-profit sent participants the UCT amounts that aligned with their treatment group assignment and that the cash groups received 87-89% of those payments in their external accounts. See Appendix Figure G.3 and Appendix Section E.3. Section 4.2.1 shows that our primary results are similar when conducting an analysis akin to a treatment-on-the-treated.

regular monthly financial flows (e.g., rent payments or welfare benefit receipts) were always at the same time relative to the survey. Appendix Table H.1 shows completion rates for the post-treatment surveys (columns “t2” through “t4”) by wave and treatment group, for a total of 16,747 survey responses after cleaning (see Appendix Section B.2 for details on the data cleaning). Participants were invited to complete all surveys; thus, non-response to an earlier post-treatment survey (e.g., t2) does not necessarily mean that participants did not respond to a later post-treatment survey (e.g., t3 or t4).¹⁵ Participants received the surveys by email and completed them online. They typically had eight days to complete the survey, though these windows were extended at times, for instance when the deadline was on a weekend or public holiday. Survey completion payments were made within a few weeks of the survey window closing.¹⁶

The four surveys had substantial, though not complete, overlap in terms of content.¹⁷ Each survey included the same questions on participants’ financial, psychological, cognitive capacity, and health outcomes; following Anderson (2008), we constructed an index for each of these four categories. The variables within the indices were standardized and weighted, with higher values being “better” (i.e., indicating higher participant well-being).¹⁸

Specifically, the financial index is constructed from data on savings stock, employment (Baird, McKenzie and Özler, 2018), work performance (if employed) (Kaur et al., 2021), work satisfaction (if employed; Leana and Meuris (2015)), earned income, subjective financial well-being (e.g., whether the participant felt behind on their finances; CFPB (2017b)), and liquidity constraints (plausibility of securing \$500 in the next three days; WorldBank (2015); Carvalho, Olafsson and Silverman (2024)).^{19,20}

The psychological index captures the participant’s sense of agency (Lachman and Weaver, 1998); the extent to which they felt they were “living their best life” (Cantril’s ladder; Kahneman and

¹⁵On average, participants took 2.0 out of the three post-treatment surveys they received, with the Control group being less responsive than the cash groups. In Section 4.6 and Appendix Section C, we discuss the correlates of responsiveness and detail the potential role of attrition in explaining the results.

¹⁶Due to an implementation error, participants in the sixth wave received the UCT payment late, when the t2 survey window was already underway. To ensure that we could still measure the effects of cash one week after receipt, participants in this wave were sent a second t2 survey a week after they received the UCT (column “t2b” in Appendix Table H.1). This second t2 survey was identical to the first t2 survey, though it was administered at a different time of month than the regular monthly schedule. In the main analyses, we ignore data from the first t2 survey unless otherwise specified, as most people who took it did so before that wave received its UCT. Instead, we use data from the second t2 survey. While unintended, this implementation error provides some natural (plausibly quasi-random) variation in survey and intervention timing, which we leverage in a later analysis (Section 4.5).

¹⁷See the Social Science Registry (socialscienceregistry.org/trials/6149) for the study materials.

¹⁸Index items that are highly correlated with other items in the index receive relatively little weight, while items that are not highly correlated (and thus contain additional information) receive comparatively more weight. See Appendix Section B.1 for details on the index construction and a table summarizing the index survey items. Note that some survey items could reasonably be categorized in multiple indices. For instance, food security is categorized in the physical health index but could also be categorized as a financial well-being measure. Our results are robust to a range of alternative re-categorizations.

¹⁹Our preanalysis plan specified using 95% winsorization on top values for unbounded variables, which include savings, earned and unearned income, and debt. In our data, this is identical to using 90% winsorized values on the top and bottom for these variables. Throughout the manuscript, we report results using these winsorizations.

²⁰Self-reported measures are prevalent in this literature, and our survey results come with the usual caveats. However, our administrative bank data allows us to partly validate at least the savings and income measures; see Appendix Section E.2.

Deaton (2010)); positive mental health (e.g., life satisfaction, how carefree they felt; Lukat et al. (2016)); how happy, anxious, and lonely they felt; and how depressed they felt (PHQ-9; Kroenke, Spitzer and Williams (2001)).

The cognitive capacity index is composed of three measures (Shah, Mullainathan and Shafir, 2012; Mani et al., 2013; Mullainathan and Shafir, 2013; Shah et al., 2018). The first is the participant’s score on nine Raven’s standard short-form matrices (Bilker et al., 2012), a measure of fluid intelligence. The second is the participant’s sense of their own “everyday memory” (Royle and Lincoln, 2008). The third measure evaluates the extent to which the participant thought about money—i.e., had “money on the mind” (Shah et al., 2018). For this measure, participants read two hypothetical vignettes that were plausibly, but not necessarily, related to money, then rated the extent to which they thought about money.²¹

The health index is constructed of participants’ self-reported general physical health, sleep quality, food security (United States Department of Agriculture, 2012), nutrition (Gallup, 2017), and exercise (Gallup, 2017; Giurge, Whillans and West, 2020).

In addition, we included various exploratory measures, including time and risk preferences (Falk et al., 2022), social well-being (Gallup, 2017), relationship with one’s partner or spouse, and self-assessed parenting quality (FragileFamilies, 2011). We describe the results of these exploratory measures in detail only when they shed additional light on potential mechanisms behind our main results. However, Appendix Figures G.8 and G.9 illustrate the key outcomes.

2.4 Financial data

Our survey data is complemented by several rich sources of financial administrative data. The first dataset allows us to observe 23,357 payments the non-profit organization sent to all 5,243 participants. These payments include UCTs, survey payments, bonuses related to our study (e.g., the lottery earnings), and occasionally payments unrelated to our study.²² When participants chose to “pull” money from their online accounts and indicated what they intended to use this money for, we observe these responses (8,438 responses).

We also observe whether participants received the money in their external accounts (17,646 attempted receipts across 5,135 participants). The vast majority (99.4%) of the attempted receipts were successful. For each successful money receipt, we observe the date, amount, and the money’s final destination—that is, whether the participant chose to receive it in their bank account (95.9%), a virtual payment card (2.2%), or a physical payment card (1.9%). See Appendix Section B.3 for more details on the sending and receiving money datasets.

In addition, participants were invited to provide access to their bank account data.²³ Forty-

²¹For instance, one of the vignettes describes a scene in which the participant needs to take an unexpected cab ride. They are asked to what extent they would have a range of non-financial thoughts while taking the ride (e.g., “Should I have tried running instead?” “It’s nice to sit back and enjoy the scene”), as well as one financial thought (“How much will this unexpected cab ride cost me?”). The vignettes varied for each survey.

²²About 2.8% of the payments made during the trial period were unrelated to our study and were instead the result of the non-profit’s normal operations.

²³Participants were paid \$10 for each account linked and were promised an increased probability of receiving cash

three percent of participants ($N = 2,261$) provided data for one or more bank accounts with data observable during the study period. About 79% of these accounts are checking accounts, 10% are savings accounts, and 10% are Paypal accounts. These data capture two key outcomes. First, they show bank account balances, typically as one “snapshot” per bank account per day (after cleaning, we observe 357,134 bank-account-balance-days). Second, they show all transactions from the account, including both credits and debits (after cleaning, we observe 850,396 total transactions). For each transaction, we observe the amount, date, category (e.g., “Food and Drink,” “Health-care,” “Bank fees”),²⁴ and a more detailed description (e.g., “McDonald’s,” “Kids Dental Place,” “Overdraft fee”). Our main analyses use 90% winsorized values (top and bottom) for the bank data, but results are robust to 95% winsorization (top and bottom). For details about these datasets, cleaning procedures, and data preparation, see Appendix Sections B.4–B.7.

3 Participants and evidence from financial data

3.1 Participants

As Table 1 demonstrates, we achieved balance across treatment groups for all the main demographics. The participant sample is majority female, majority non-White (74% of participants who identify as not exclusively White identify as Black or African-American), majority high school graduate, majority parents, and majority without a spouse or partner. Most participants lived in urban areas. Together, they represented 45 states and the District of Columbia.

Most participants were living in poverty. Our calculations indicate that about half were under the federal poverty line in 2019,²⁵ and the majority suffered additional financial strain as a result of the pandemic.²⁶ From self-reports, the median earned household income in the month before the t1 survey was \$414 (average: \$881), median total household income was \$1,028 (i.e., earned plus unearned income²⁷; average: \$1,455), and debt stock was on average 44 times larger than savings stock. Median savings reports were \$0, and 81% reported having under \$100. These numbers are largely corroborated by the bank account data. The median sum of all bank inflows (which can proxy for income) over the 30 days before t1 was \$1,126.19 (average: \$1,552.23); see Appendix Section E.2 and Appendix Figure G.11. Median bank balances at t1 were \$0.86 (average: \$114.45).

Table 1 also displays baseline (pre-randomization) index values for each group. We achieved

transfers after the study was over. The likelihood of providing data for at least one bank account is positively associated with being in the \$500 or \$2,000 group, answering more post-treatment surveys, being younger, being a parent, and having a lower baseline financial index score. Baseline psychological index, cognitive capacity index, and health index scores do not predict the likelihood of providing bank data. See Appendix Table H.2.

²⁴These categories were generated by the financial services company that provided the data.

²⁵Calculated using the Health & Human Services poverty guidelines (<https://aspe.hhs.gov/2019-poverty-guidelines>; accessed 19 April 2023) using participants’ 2019 household income rounded up to the nearest \$10,000 and household size measured in 2020.

²⁶While 30% of participants were unemployed before the pandemic began, this number rose to 57% by the time of the t1 survey. Nearly three-quarters of participants reported losing their job, having their work hours involuntarily reduced, and/or losing business income because of the pandemic.

²⁷Sources of unearned income include public assistance or welfare payments, governmental stimulus checks, tax refunds, non-profit organization payments, unemployment benefits, child support, and retirement income.

Participant Demographics and Baseline Index Values.

	Control	\$500 Group	\$2,000 Group	F	p-value
% Female	86	86	87	0.09	0.915
Age	35	35	36	1.68	0.187
% Non-White	77	80	78	2.78	0.062
% More than high school	59	61	60	0.43	0.654
Household size	3.8	3.7	3.7	2.99	0.051
% Parent	82	80	82	1.41	0.243
% Married/partner	43	42	44	0.32	0.728
% Employed	43	43	45	0.93	0.394
Savings stock (\$)	405	470	373	1.87	0.155
Debt stock (\$)	18,750	18,259	17,513	0.59	0.554
Earned income last mo. (\$)	859	910	920	1.52	0.219
Unearned income last mo. (\$)	517	535	517	0.34	0.713
% Under FPL 2019	51	49	49	1.05	0.351
Financial index at t1	0	0.055	-0.008	1.58	0.206
Psychological index at t1	0	0.054	0.013	1.39	0.248
Cognitive capacity index at t1	0	0.053	0.009	1.37	0.253
Health index at t1	0	0.079	0.054	3.25	0.039

Table 1: Means for each treatment group. The F test statistic and corresponding p -value refer to a one-way ANOVA testing for the differences across treatment groups. All variables were measured in t1. FPL=Federal poverty line. See Section 2.3 for a description of the indices. In the last four rows, the financial, psychological, cognitive capacity, and health indices are 0 for the Control group by construction.

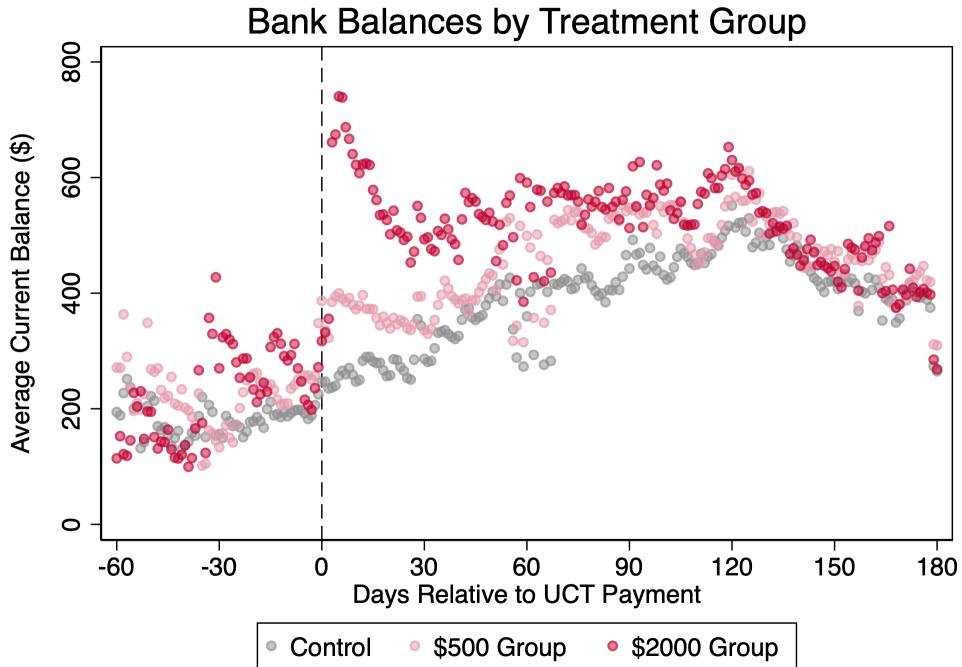


Figure 2: Bank balances by day, for participants who provided access to bank data. X-axis denotes days relative to when each participant’s wave received its UCT.

balance on the financial, psychological, and cognitive capacity indices. For the health index, the cash groups had slightly but significantly higher baseline values than Control, which—if anything—makes the negative values we observe in the cash groups in the post-treatment surveys more notable.

3.2 Bank account balances and spending

For the subset of participants who provided access to their bank accounts, we can examine whether the cash transfers can be observed in their bank account balances and, more intriguingly, how long they stay there. Figure 2 illustrates the bank account balances over time by treatment group. Because participants were enrolled across multiple waves and thus received their UCTs on different calendar dates, we plot the balances as a function of days relative to that wave’s UCT date. As expected, there were no differences in average daily bank account balances across treatment groups before the UCT date (OLS regression with treatment group indicators and robust *SEs*: intercept = \$172.11, $\beta_{\$500}$ = \$21.59, $p = 0.370$; $\beta_{\$2000}$ = \$47.06, $p = 0.145$).

Immediately after the UCT was paid, it was reflected in participants’ daily bank account balances.²⁸ In the first two weeks following payment, the \$500 group on average had \$43 more in their bank accounts, while the \$2,000 group had \$213 more (controlling for pre-UCT daily balances; see

²⁸The reason the increase in bank balances was less than \$500 for the \$500 group and less than \$2,000 for the \$2,000 group, even immediately after the transfer, is due to the fact that participants occasionally provided data access to bank accounts that were not the same accounts as those into which they received their UCTs. We are thus averaging in a number of zeroes.

Appendix Table H.3). However, the differences were short-lived. By the second two-week period, there was no statistically significant difference between the \$500 group and Control; by the third two-week period, there was also no difference between the \$2,000 group and Control.²⁹

What explains the rapid closing of the gap between treatment and Control group bank account balances? One possibility is that UCT recipients spent the money; another is that they transferred it to bank accounts we do not observe. To disentangle these explanations, we turn to the financial transactions data.³⁰ First, we find that while the Control group spent an average of \$68.70 per day in the first two weeks following the UCT payment date (not controlling for pre-UCT spending), the \$500 group spent an additional \$26.33 per day ($p < 0.001$) and the \$2,000 group spent an additional \$81.67 per day ($p < 0.001$). That is, the \$500 and \$2,000 UCTs resulted in the recipients spending 138% and 219% of what would have been their regular daily spending amounts, respectively. These numbers correspond to a 74% and 57% marginal propensity to consume out of the \$500 and \$2,000 windfalls, respectively, just in the first two weeks following cash receipt.³¹ By the second two-week period, however, the difference in spending drops substantially and loses significance (see Figure 3 and Appendix Table H.4).³² We observe a similar pattern for transaction volume, debit volume, and net expenditures (debits minus credits); see Appendix Section E.4.³³

Because we observe only 43% of the sample during the trial period, and the sample is self-selected, it is important to question whether these results also apply to the unobserved participants. When we restrict the sample to just the 43% and run our primary analyses on the survey outcomes, we find that our conclusions about the survey outcomes are virtually unchanged. These results are consistent with the notion that the bank balance and financial transaction results would generalize to the remainder of the sample.

Given that most participants seem to have spent their UCT money within a couple of weeks of receiving it and our earliest survey was one week after cash receipt, it follows that most participants responded to the survey after they had already spent some (potentially substantial) portion of

²⁹Results are robust to controlling for the number of days between Wave 1's measurement and the participant's measurement instead of wave number, as well as to controlling for bank account type (e.g., depository, loan, credit) and subtype (e.g., checking, savings). Without controlling for average pre-UCT balance, the \$2,000 group retains a significantly higher balance than Control through the fourth fortnight.

³⁰To ensure we capture the effects of the treatment and not the treatment itself, for these analyses we exclude payments from the non-profit organization that occurred on or after the trial start date (4,328 payments total).

³¹See Mankiw (2000) for a model of how people who live paycheck-to-paycheck may respond differently to fiscal policies, relative to people who save. Empirical work has shown that marginal propensity to consume tends to be higher among low- (vs. high-) income recipients of cash transfers (Johnson, Parker and Souleles, 2006; Ruggeri et al., 2022; Baker et al., 2023; Chetty et al., 2024; Golosov et al., 2024). See Egger et al. (2022) and Karger and Rajan (2020) for similar marginal propensity to consume figures among low-income UCT recipients.

³²Results are robust to controlling for the number of days between the Wave 1 measurement and the participant's wave's measurement instead of wave number, controlling for bank account type and subtype, and not controlling for pre-UCT spending.

³³We can also test if post-treatment *inflows* into bank accounts vary by treatment group (excluding payments from the non-profit). If they do, this could be a sign that the UCT shifted participants' likelihood of working for pay, seeking repayment for loans outstanding, seeking additional governmental or non-profit support, selling possessions, etc. Among the sample that provided bank data access, we find that both treatment groups have somewhat elevated inflows after the UCT, but these differences are not significant (OLS regressing average inflows per day following the UCT on treatment group dummies, robust SEs: intercept=\$93.37; $\beta_{\$500} = \$8.45, p = 0.059$; $\beta_{\$2000} = \$5.32, p = 0.333$).

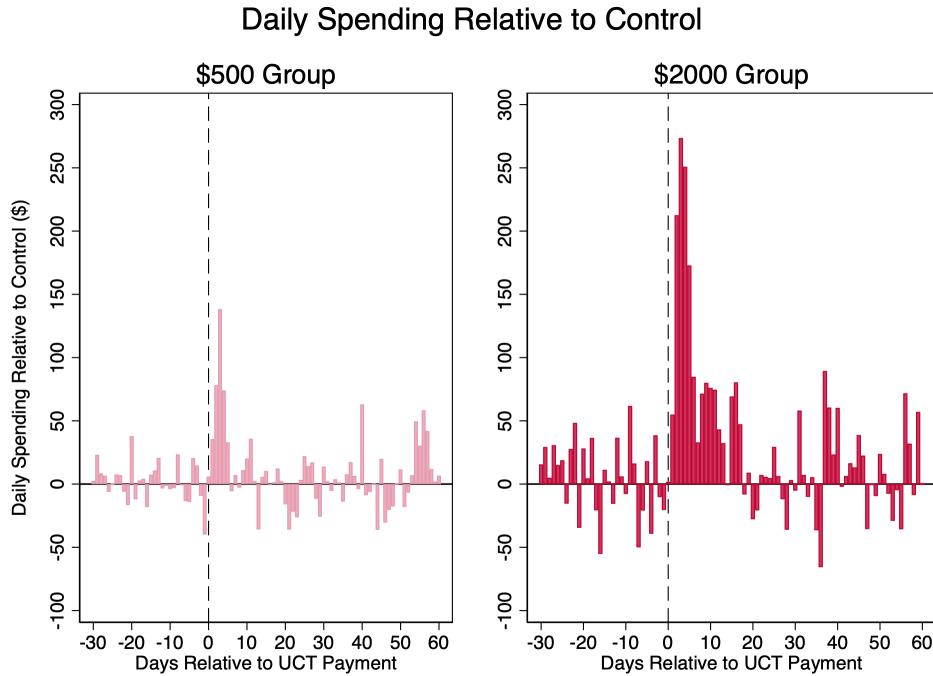


Figure 3: Daily spending of the \$500 group (left) and \$2,000 group (right), relative to Control, for participants who provided access to bank data. Positive values indicate that the given group spent more than Control. X-axis denotes days relative to when the participant’s wave received its UCT.

their money. This leaves open the possibility that even our earliest survey failed to capture some immediate effects of cash. To test this, we leverage the natural variation in Wave 6 timing to examine survey responses only a few days after cash receipt; see Section 4.5.

3.2.1 Spending patterns

What did participants spend the UCT money on? To answer this question, we use two data sources.

First, recall that participants who withdrew money from the online platform (whether the UCTs, study participation payments, or both) were asked how they intended to use the money they were withdrawing. Two research assistants blind to the hypotheses and treatment groups coded each of these responses (8,438 responses across 3,331 participants) (see Appendix Section B.8 for details). Figure 4 illustrates those codes for each treatment group. The cash groups were most likely to report intending to spend the money on general “bills” (which could not be further categorized into more specific bills, such as utilities or credit cards), groceries, and transportation, and were far more likely than the Control group to intend to use the money for bills and housing. This data has two major advantages. First, it covers the majority of the participants. Second, it provides insights into how participants may have earmarked the money for particular expenditures (Thaler, 1999). However, it also has some disadvantages. Participants may have felt uncomfortable reporting their true intended usage (Godoy, Karlan and Zinman, 2021), or they may have intended

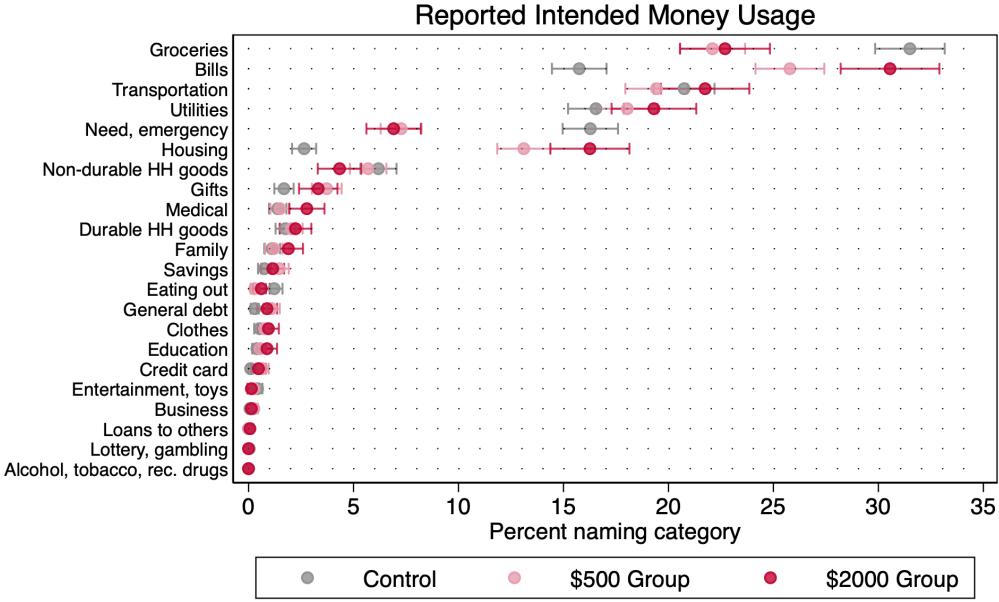


Figure 4: Participants’ reported intentions of how they would use the money they were withdrawing. Each response could be coded as multiple categories. Excludes the “Other” category and missing values. HH=household, rec. drugs=recreational drugs. Circle denotes mean value, bars denote 95% CIs.

to use the money for one thing and ultimately used it for another. Because of this, we also analyze a second—more objective—dataset: the financial transactions data.

To analyze this data, we take the categories the financial services company generated to describe each transaction and regress daily spending in each of those categories (from the UCT date to the final day of the t4 survey) on treatment group dummies. Consistent with prior work on economic stimulus check spending (Misra, Singh and Zhang, 2022), the results reveal that much of the UCT money was spent on “transfers,” a fairly broad category that includes digital payment vehicles (e.g., Venmo, Paypal, Zelle), ATM withdrawals, and loading of pre-paid debit cards. Cash groups also spent significantly more on shops (e.g., Dollar Tree, Amazon), food and drink (grocery stores, restaurants), and travel (e.g., gas, parking fees) (see Figure 5). We do not detect differences in the amount of bank fees charged (Stango and Zinman, 2014).³⁴ These results are robust to including a range of covariates, including daily spending before the UCT. When restricting to just the first two weeks following the UCT payment, results are quite similar, although somewhat more poorly specified, given the rarity of transactions in certain categories (e.g., healthcare, taxes).

³⁴This is also true when restricting to what we call “bad” bank fees (insufficient fund fees, overdraft charges, cash advance fees, late payment fees, and generic bank fees—as opposed to bank fees that are less likely to be an indication of financial stress, such as ATM withdrawal fees and foreign transaction fees). This may be because some banks waived certain fees during COVID (Toh and Tran, 2020) and the overall incidence of these fees is fairly low in our dataset.

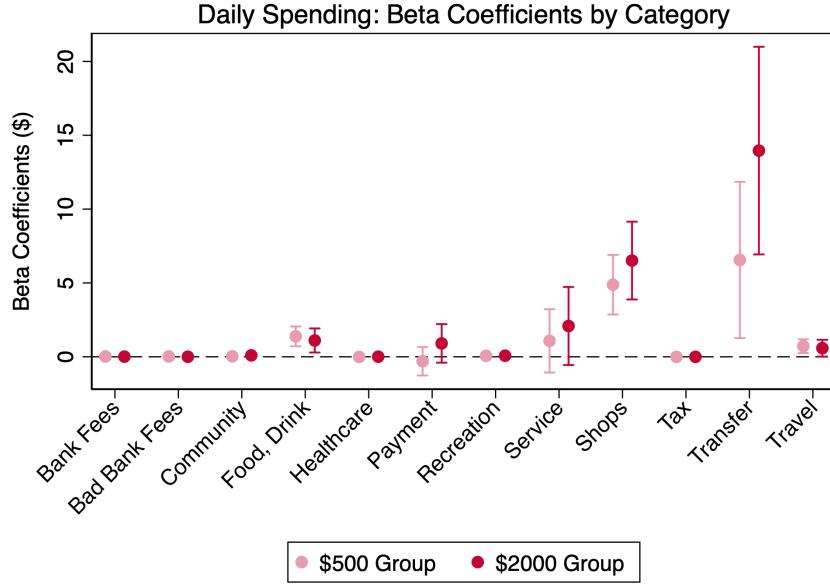


Figure 5: How much more the cash groups spent per day relative to Control, by spending category, for participants who provided access to bank data. Beta coefficients are from regressing average daily spending over the trial period (between when the participant’s wave received its UCT and the final day of the final survey) on treatment group indicators with robust *SEs*. Error bars denote 95% CIs. See text for description of the “bad” bank fees category.

4 Evidence from survey responses

4.1 Analytical approach

We now turn to our preregistered survey outcomes. Our identification strategy is based on random assignment to treatment group and we use an intent-to-treat approach in our primary analyses. As described in Section 2.3, we construct summary indices (Anderson, 2008) for each of four categories: financial, psychological, cognitive capacity, and health outcomes. By construction, the Control group average is zero at each time point. Our primary specification to estimate treatment effects is:

$$y_{i,t>1} = \beta_0 + \beta_1 500_i + \beta_2 2000_i + \beta_3 OP_i + \delta y_{i,t=1} + \epsilon_i \quad (1)$$

where y is one of four composite indices (financial, psychological, cognitive capacity, and health) for individual i at time t and $t = 2, 3$, and 4 are the post-treatment surveys. “500” and “2000” are indicator variables that equal 1 if the participant was in the \$500 or \$2,000 group, respectively. The omitted category is the Control group. OP is an indicator variable that equals 1 if the participant had access to the additional non-financial components of the non-profit organization’s online platform (see Section 2.2). Finally, $y_{i,t=1}$ is the baseline measure of the composite index, included to improve statistical power (McKenzie, 2012).³⁵ In our primary analysis, we collapse

³⁵An exception is the cognitive capacity index. Because we did not administer Raven’s matrices at baseline (by design, to avoid a participant becoming more familiar with, and perhaps learning how to efficiently solve, the puzzles

across post-treatment time periods (i.e., we structure the data such that each row corresponds to a participant-time period response, with up to three post-treatment rows per participant) and cluster robust standard errors at the participant level.

Given the number of parameters we test (in our primary analyses, there are eight parameters: 2 treatment groups \times 4 indices), we conduct multiple hypothesis testing corrections. We do this in two ways. First, we control the false discovery rate using the Benjamini-Hochberg approach (henceforth, “BH”) (Simes, 1986; Benjamini and Hochberg, 1995), which can be used for both independent (Benjamini and Hochberg, 1995) and positively correlated tests (Benjamini and Yekutieli, 2001). Second, for comparison, we control the family-wise error rate, where the “family” of statistical tests are the eight parameters mentioned above. In particular, we use the Westfall-Young approach (henceforth, “WY”; Westfall and Young (1993); Westfall and Troendle (2008); Jones, Molitor and Reif (2019)), which uses bootstrap sampling to allow for dependence across outcomes.

We conduct BH and WY corrections for our preregistered survey outcome analyses (Section 4.2) and each robustness check (Section 4.2.1). The WY corrections result in fewer hypotheses being rejected and the prespecified BH corrections yield very similar rejection conclusions as the unadjusted p -values. Regardless, the primary conclusions are fairly similar with or without either set of corrections. Thus, for brevity and consistency across the rest of the manuscript, for each analysis we run, we report the unadjusted p -values in the text. For the few cases in which the unadjusted p -values reach significance at a standard $\alpha = 0.05$ level but the WY- and/or BH-adjusted values do not, we indicate this in the text.³⁶ Appendix Table H.5 reports the unadjusted, BH adjusted, and WY adjusted p -values for each of our analyses when we aggregate across post-treatment time periods (and thus have eight β coefficients). Appendix Table H.6 reports the three types of p -values when we disaggregate by post-treatment time period (and thus have 2 treatment groups \times 4 indices \times 3 post-treatment time periods = 24 β coefficients).

4.2 Survey results

The main results of the survey outcomes using our prespecified analyses can be summarized as follows: Against our expectations (and the expectations of laypeople and experts, as discussed in Appendix Section A), we find no evidence that (more) cash had a positive effect on self-reported survey outcomes for any of our predetermined specifications. In fact, both cash groups reported experiencing worse outcomes than the Control group on the financial ($\beta_{\$500} = -0.096, p < 0.001$; $\beta_{\$2000} = -0.058, p = 0.047$, not significant with WY correction), psychological ($\beta_{\$500} = -0.109, p < 0.001$; $\beta_{\$2000} = -0.130, p < 0.001$), and health indices ($\beta_{\$500} = -0.122, p < 0.001$; $\beta_{\$2000} = -0.143, p < 0.001$). On the cognitive capacity index, there were no statistically significant differences between the Control group and the cash groups, although the coefficients were negative, as well ($\beta_{\$500} = -0.049, p = 0.092$; $\beta_{\$2000} = -0.070, p = 0.061$).

in the future), the baseline cognitive capacity index is constructed using only the other variables in this index.

³⁶There are a few cases where the BH adjusted values reach significance but the unadjusted values do not. We do not flag these in the text, but all values are available in Appendix Tables H.5 and H.6.

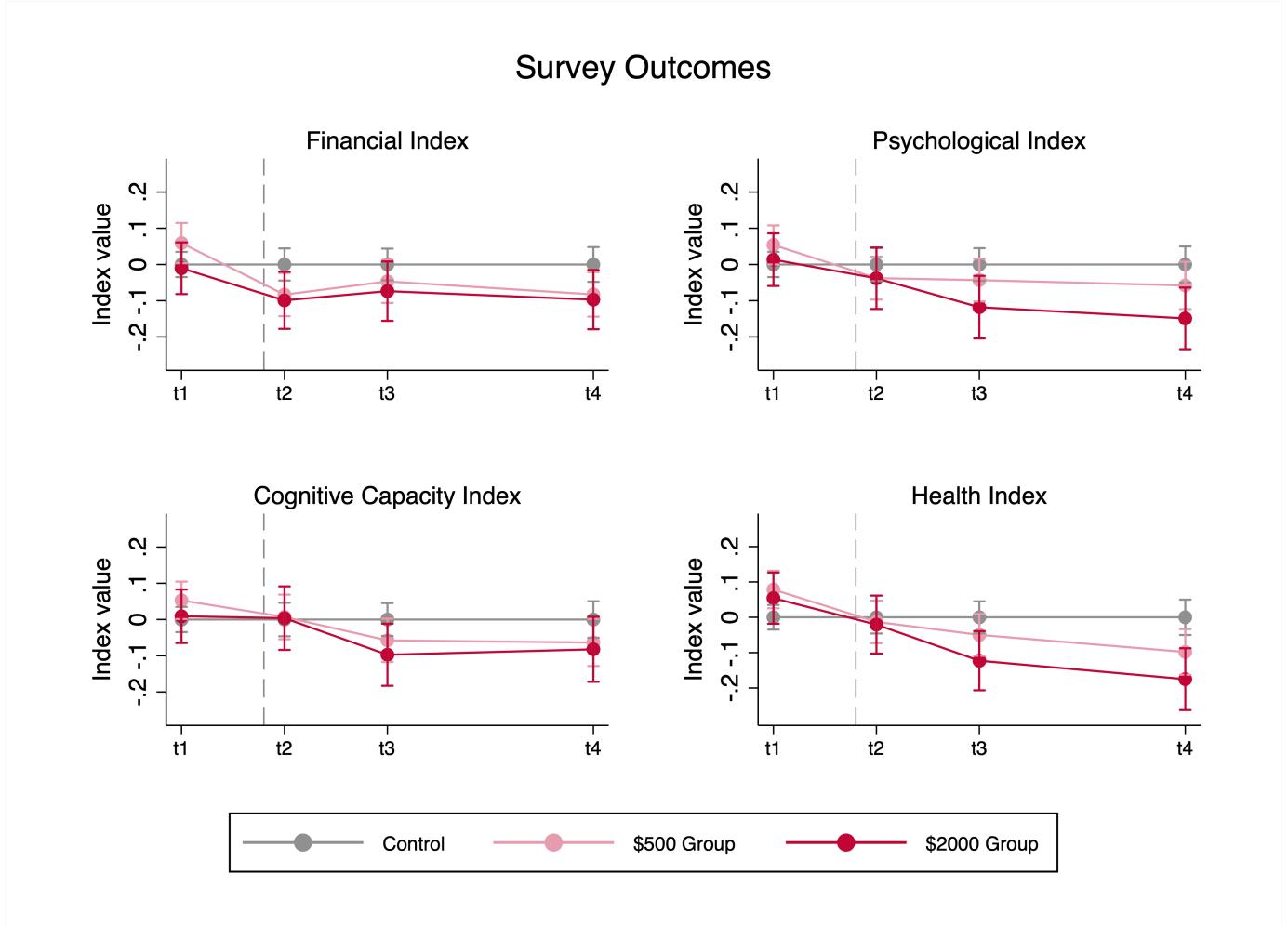


Figure 6: Main survey outcomes for the four prespecified indices. Vertical dashed line indicates intervention; t1 was the baseline survey before randomization and before receiving cash. Error bars denote 95% CIs.

There were no statistically significant differences between the two cash groups for any of the indices (all $p \geq 0.228$).

Figure 6 and Table 2 summarize our main results. Appendix Figures G.4 through G.7 plot the results by individual variable (standardized as Z-scores but not weighted), visualizing both treatment group differences and how variable values changed over time. As explained in Section 4.6, attrition could have played some role in making the effects appear more negative than they might have otherwise been; however, we also show that positive effects are highly unlikely. We therefore refer to our results as “non-positive.”

Indeed, the effect sizes we do observe are non-negligible. Appendix Section E.5 offers a back-of-the-envelope benchmarking exercise to better understand the size of these index effects. Compared to losing one’s job due to the pandemic (up to 14 months earlier), the relative impact of the UCTs in our study ranged from about one-quarter for the financial and psychological indices, to two-thirds for the health index, to unity for the cognitive capacity index.

Effect of UCTs on Survey Indices.

	(1) Fin.	(2) Psych.	(3) Cog. Cap.	(4) Health
\$500 Group	-0.096 (0.022)	-0.109 (0.024)	-0.049 (0.029)	-0.122 (0.024)
\$2,000 Group	-0.058 (0.029)	-0.130 (0.031)	-0.070 (0.037)	-0.143 (0.030)
Fin. Index at t1	0.635 (0.012)			
Psych. Index at t1		0.623 (0.011)		
Cog. Cap. Index at t1			0.294 (0.013)	
Health Index at t1				0.611 (0.012)
Online Platform	-0.014 (0.020)	-0.025 (0.021)	-0.024 (0.026)	-0.012 (0.021)
Constant	0.016 (0.017)	0.043 (0.018)	0.019 (0.022)	0.030 (0.018)
Observations	10271	9774	9582	9704
R^2	0.415	0.399	0.087	0.382

Table 2: OLS regressions. Collapsing across all post-treatment time points. Standard errors (in parentheses) are clustered at the participant level and robust. “Online Platform” is a binary indicator for having access to additional features of the online platform.

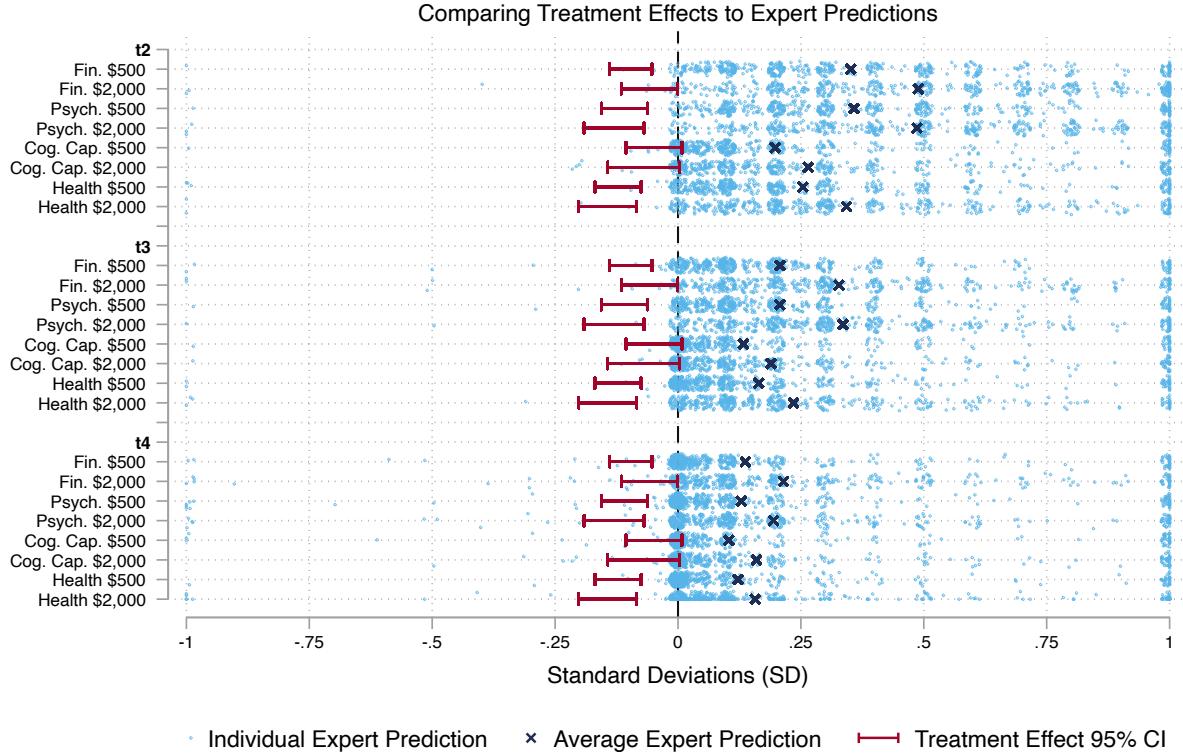


Figure 7: Comparison of the observed, preregistered treatment effects in the RCT with the incentivized predictions of experts in the prediction survey.

It is also instructive to compare the observed treatment effects with the results of the prediction study, which captured experts' and a nationally representative sample's prior expectations about the effects of these treatments (for details, see Appendix Section A). Figure 7 shows how the 95% CIs of the preregistered RCT analyses compare to experts' predictions. The figure illustrates that many experts were too optimistic about the effectiveness of such cash transfers in this context, particularly about the short-run effects of the treatment (see t2 measurements, i.e., one week after the treatment). In fact, across all time periods and indices, very few experts predicted any negative effects. The equivalent predictions of the nationally representative sample—not plotted here—are even more optimistic; see Appendix Section A.2. The mismatch between predictions and the reality of the treatment effects holds when taking into account attrition; see Section 4.6.

4.2.1 Robustness

The non-positive effects of cash on the survey outcomes are robust to a wide range of checks; see Appendix Tables H.5 and H.6 for details of each robustness check. We summarize the key conclusions here.

In a first robustness check, we restrict our analysis to only participants who, in the t2 survey, were able to correctly identify how much cash they had received from the non-profit in the previous

month ($N_{Control} = 1,951$; $N_{\$500} = 650$; $N_{\$2000} = 292$). This allows us to conduct an analysis akin to a treatment-on-the-treated approach, examining the effects among participants who were aware of the UCT and its source. It also allows us to examine only those participants who were relatively high in attention to and comprehension of the survey. Rerunning the analyses described in Section 4.1, we find that the results are similar, though the effect sizes become less negative and in some cases are not distinguishable from 0. Again, there are no differences between the cash groups (all $p \geq 0.397$). Results are similar when controlling for wave number and a range of demographics and financial variables measured at t1.

We next restrict the data a second time, only retaining people who both correctly identified how much cash they received and answered all four surveys ($N_{Control} = 1,148$; $N_{\$500} = 485$; $N_{\$2000} = 215$). With this analysis, most of the point estimates continue to be negative but largely lose significance, as would be expected given the substantial drop in sample size. Yet again, there were no differences between the two cash groups (all $p \geq 0.369$).

In addition, our conclusions are qualitatively similar to those of the main specification when we control for a range of additional covariates (wave number, gender, age, race, education, household size, parent status, partner status, employment at t1, savings at t1, debt at t1, last month's earned income at t1, last month's unearned income at t1, and a binary indicator for being under the federal poverty line in 2019), drop Wave 6 entirely, and use Z-scored indices instead of the Anderson (2008) constructions. In all cases, neither treatment resulted in positive effects for any index; see Appendix Table H.5. When running the analysis described in Section 4.1 but not controlling for the relevant baseline index value, we find that eight of eight treatment β values are still negative (though generally closer to 0 than when we control for the baseline value), and half are statistically significantly different from 0. When splitting the data by time period rather than examining all post-treatment values together, we overwhelmingly still see negative effects but reject the null less often; see Appendix Table H.6.

In a final robustness exercise, we disaggregate the results by variable and time, examining each of the prespecified variables separately at each time point rather than weighting them and placing them into indices. Specifically, we regress each of the 40 prespecified variables (oriented such that higher values are better) on dummies for the treatment groups separately at each time point. Of the 240 β coefficients (40 variables \times 2 treatment dummies \times 3 post-treatment time periods), 142 are not significantly different from 0, two are statistically significantly positive (using $\alpha = 0.05$), and 96 are statistically significantly negative. This analysis suggests that the results observed at the level of aggregated indices were not driven by only a few outlier variables.

In summary, across a range of robustness checks, we find no evidence that cash improved the participants' financial, psychological, cognitive capacity, or health measures. Across most specifications, the cash groups appeared worse than the Control group on the psychological and health outcomes. In some instances, they also appeared worse on financial outcomes. They typically did not differ significantly in the cognitive capacity outcomes.

4.3 How did receiving cash make participants feel?

To better understand how cash affected participants’ life events—and, importantly, their experiences of those life events—we explore whether there were systematic differences in how participants responded to more objective versus more subjective survey questions (Ackerman and Paolucci, 1983; Perrig-Chiello, Perrig and Stähelin, 1999; Oswald and Wu, 2010; CFPB, 2017a).

To this end, we conduct the following exercise (see Appendix Section E.6 for additional details): We first take all the survey questions—for completeness, we include all survey items, regardless of whether they were in the prespecified indices—where one end of the response scale would generally unambiguously indicate higher participant well-being (e.g., we include happiness and health, but do not include risk or time preferences). We then orient them such that higher values are better. Two independent coders then categorized each question as being more “objective” or “subjective.” Objective questions measure quantifiable or countable outcomes that could in theory be verified if data were available (e.g., housing status, the number of days that an event occurred), while subjective questions capture how a participant felt about or experienced something in their life (e.g., how they rated their sleep quality in the past week, how anxious or stressed they felt).

We conduct two analyses: the first focuses on effect size; the second, on significance testing. In the first analysis, we create separate indices for the objective and subjective variables. We then run two regressions, first regressing the objective index on treatment group dummies (examining all post-treatment values and using robust *SEs* clustered at the participant level), and then repeating this analysis for the subjective index.

In the second analysis, we examine each variable individually instead of using an average for the two categories. Specifically, we regress each of the objective and subjective survey variables on binary indicators for each treatment group (again examining each measure at all post-treatment time periods and using robust *SEs* clustered at the participant level). For each regression, we count the number of treatment group dummy variable coefficients that (a) indicate that the cash group is doing worse than Control; and (b) are statistically significant at the $\alpha = 0.05$ level.

The two analyses converge to the same conclusion: the negative effects are concentrated among the subjective outcomes. When examining the objective outcomes, the effects cannot be distinguished from 0. See Figure 8. Coupled with the bank account and spending results, we interpret these data as providing suggestive evidence that the cash had no effect (or a positive effect, if one considers the increase in bank account balances and spending) on the more objective outcomes, but a more negative effect on the more experienced, subjective outcomes. That is, although cash made people better off—or at least no worse off—objectively, it made them *feel* relatively worse off.

4.4 Can participants’ characteristics explain more positive or negative effects?

We explore potential heterogeneity across participants by re-running the primary analyses described in Section 4.1, each time interacting the treatment group dummies with a potential moderator variable. We explore the following moderators: the cost of living in the participant’s zipcode, receiving a governmental stimulus check in the past month, receiving governmental unemployment

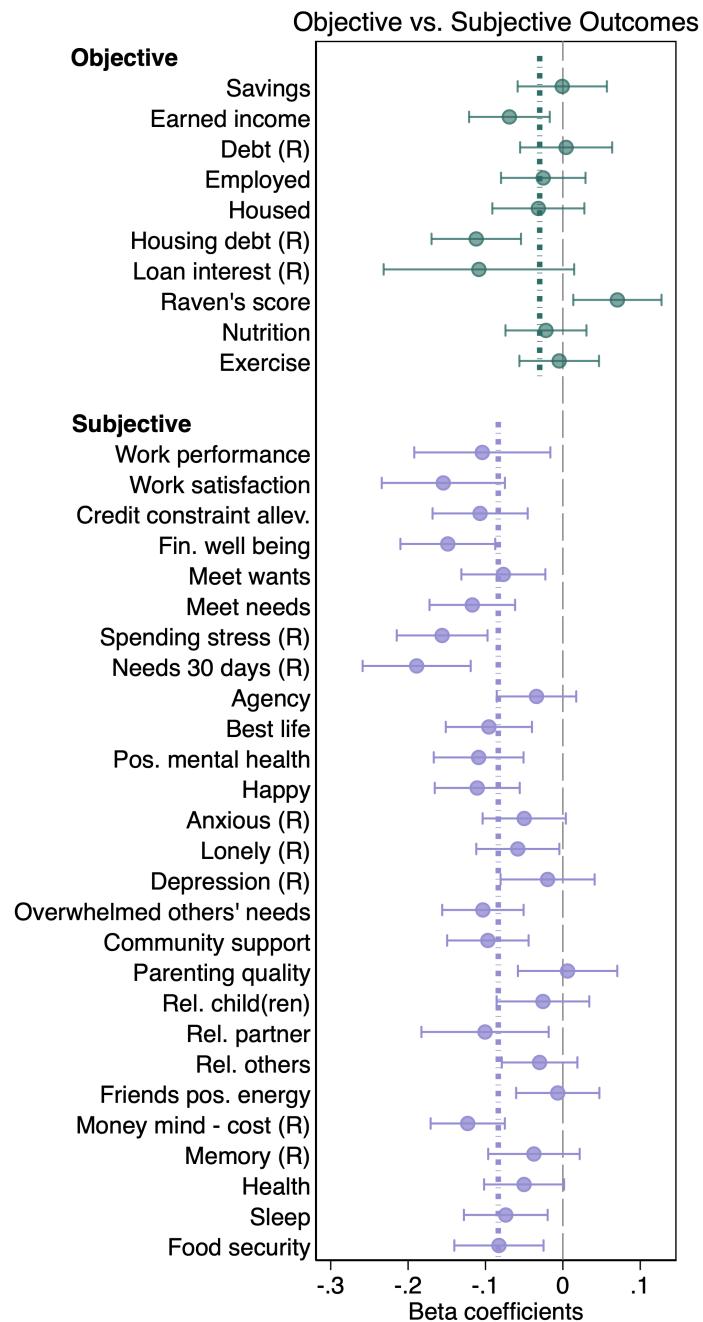


Figure 8: Beta coefficients from separately regressing each survey construct written on the y-axis on a binary indicator for being in a cash group, collapsing across all post-treatment surveys and using robust standard errors clustered at the participant level. The omitted category is the Control group. Each construct is standardized into a Z-score to ensure comparability. Constructs denoted with “(R)” contain only variables that were reverse coded from the original survey data, such that higher values are always better in this graph. Includes all participants who responded to at least one survey question in a post-treatment survey. Error bars denote 95% CIs. The thicker vertical dashed lines in the objective and subjective sections represent the means of the β coefficients for those sections.

benefits in the past month, being employed immediately before the pandemic, being employed at t1, unearned income at t1, earned income at t1, credit constraints at t1, savings at t1, financial well-being at t1, being under the federal poverty line in 2019, being female, being not exclusively White, age, completing at least high school, being a parent, number of children, the fraction of one's children who are male, having a partner or spouse, and patience at t1. Of the 160 β coefficients tested (20 moderators \times 2 treatment group dummies \times 4 index values), only five reach statistical significance at the $\alpha = 0.05$ level.³⁷ Given the obvious multiple hypothesis testing concern, the lack of robustness, and the lack of a coherent story tying the results together, we conclude that there is no compelling evidence for participant-level heterogeneity within this sample.

4.5 Are potentially positive effects simply short-lived?

The panel survey data allows us to test for fine-grained changes in how the UCTs affected outcomes over time. First, analyses reveal that the effects of cash appear relatively stable across our three post-treatment time periods for the financial and psychological indices (all $AnyCash \times t$ dummies: $p \geq 0.138$). Cash appears to produce weakly more negative effects over time for the cognitive capacity index (relative to $AnyCash \times t = 2$: $\beta_{AnyCash \times t=3} = -0.077, p = 0.044$, $\beta_{AnyCash \times t=4} = -0.076, p = 0.053$) and health index ($\beta_{AnyCash \times t=3} = -0.058, p = 0.061$, $\beta_{AnyCash \times t=4} = -0.109, p = 0.002$).

As shown in Appendix Table H.6, we find no evidence of cash improving outcomes even when we restrict our analyses to just t2 measurements, i.e., one week after UCT payment. Of course, it is possible that cash had even shorter-lived effects in the first few days after cash receipt that then dissipated or changed by the seventh day. To test this possibility, we exploit the natural variation in timing caused by the Wave 6 treatment administration error (see Footnote 16), in which some participants received their UCTs *during* a t2 survey window. We restrict the sample to just Wave 6 participants who responded to the (first) t2 survey one to six days after the UCT was sent (26% of the wave; $N_{Control} = 142$; $N_{\$500} = 107$; $N_{\$2000} = 68$). Re-running the regressions from Section 4.1, we find no statistically significant differences between the three treatment groups on any index. We similarly cannot reject the null when allowing for an extra day for the cash to arrive in the participants' accounts, and also when collapsing across the two cash groups instead of examining the \$500 and \$2,000 groups separately (all $p \geq 0.236$). These analyses should be interpreted with caution, as the people who responded to the survey several days after it was sent were a self-selected minority and analyses may be underpowered. Nevertheless, they suggest that even when looking at a shorter time frame, there is no evidence of positive effects of cash on survey outcomes.³⁸

³⁷These are: $\$500 \times Savings$ on health ($\beta = 0.0001, p = 0.006$), $\$2,000 \times NonWhite$ on cognitive capacity ($\beta = 0.200, p = 0.024$), $\$2,000 \times FractionSons$ on cognitive capacity ($\beta = 0.305, p = 0.039$), $\$2,000 \times Partner$ on cognitive capacity ($\beta = -0.159, p = 0.035$), and $\$2,000 \times Patience$ on health ($\beta = 0.046, p < 0.001$).

³⁸We also exploit the Wave 6 administration error to test the extent to which survey outcomes changed for Wave 6 participants who answered the two t2 surveys only about nine days apart with a UCT payment inbetween. We again find no evidence of cash improving outcomes; see Appendix Section E.7.

4.6 Can attrition explain the lack of positive effects?

Forty-six percent of participants responded to all three post-treatment surveys, 23% responded to two, 15% responded to one, and 17% did not respond to any. Of these 17%, about 18% provided bank account data. In total, thus, we have post-treatment survey data for 83% of our participants and other post-treatment bank data for an additional $0.17 \times 0.18 = 3$ percentage points of participants, for a total of 86%. See Appendix Table H.1. We only briefly summarize the potential role of attrition here but refer interested readers to our more extensive analyses and discussion in Appendix Section C.

Could differential attrition explain our results? Post-treatment responsiveness was lower among the Control group (80%) than the \$500 group (90%) and the \$2,000 group (88%), but there was no significant difference between the two cash groups' responsiveness. It is plausible that differential attrition across the cash and Control groups could have biased estimates of the effects of receiving any cash—relative to no cash—downwards. That is, if the unobserved Control group participants had particularly bad post-treatment outcomes and/or the unobserved cash group participants had particularly good post-treatment outcomes, the real effects of cash could be more positive than we observe.³⁹ We consider several scenarios that could produce such a pattern of results. These include variation in participants' baseline financial need, macroeconomic or public health conditions that affect survey uptake (e.g., COVID conditions or unemployment in their area), financial shocks after baseline, reciprocity or trust towards the experimenter, and differing beliefs about whether responding would yield financial support. In Appendix Section C.4, we statistically evaluate the merits of these scenarios. Overall, we find some limited evidence for the final scenario, but little to no evidence for the others.

Absent differential attrition, are there positive effects of more cash? While the Control group's survey uptake is lower, the response rates are similar and not statistically different between the two cash groups. This allows us to examine the effect of receiving more cash in the *absence* of differential attrition. As Figure 6 and Table 2 show, even quadrupling the cash amount—\$500 versus \$2,000—does not reveal positive effects. Thus, to the extent that differential attrition could explain the lack of positive effects between the Control and cash groups, no such argument can explain the lack of positive effects between the two cash groups.

Are certain types of participants more likely to attrit, leading to lack of positive effects? We examine a large set of potential predictors of responsiveness, operationalizing it in both a binary and continuous way. Importantly, Appendix Section C.1 and Appendix Tables H.7 and H.8 reveal that none of the indices at baseline predict responsiveness, nor do additional baseline financial characteristics. A few demographics—age, race, and household size—carry some predictive power. However, as shown in Section 4.2.1 and Appendix Table H.5, the results are unchanged when controlling for these and many other variables at baseline.

³⁹But note that the opposite could be true instead: if the unobserved Control group participants had particularly good outcomes and/or the unobserved cash group participants had particularly bad outcomes, the real effects of cash could be even more negative than we observe. See Appendix Section C.3 for a more detailed discussion.

Is this response rate atypical? We observe one or more post-treatment survey responses from 83% of participants. If the response rate were substantially lower than in comparable studies, it could imply that we did not hear from a specific subset of participants who would have responded in another trial, which could affect differences in conclusions. However, our response rate is comparable to many recent cash transfer trials in North America, including Dwyer et al. (2023) (52%), Liebman et al. (2022) (95%), and Yoo et al. (2022) (93%), as well as the two studies most comparable to ours, Jacob et al. (2022) (65%) and Pilkauskas et al. (2023) (42% to 61%). Thus, to the extent to which our results differ from those of other trials, it is unlikely that the differences are driven by an unusual response rate.

Are participants who respond to the surveys systematically different across treatments? We conduct a selective attrition test to identify whether, conditional on response status, observable characteristics are balanced across treatment groups (Ghanem, Hirshleifer and Ortiz-Beccera, 2023). The null is a joint hypothesis of the equality of baseline outcome distributions between respondents in the three treatment arms, as well as attritors in the three treatment arms. Regardless of whether we focus just on the baseline index values or the baseline index values plus an array of covariates collected at baseline, the tests reveal that observable characteristics are balanced (Appendix Section C.2), helping to alleviate such concerns.

Are positive effects (mechanically) possible? To identify whether our data could, in theory, support the possibility of (more) positive effects, we calculate Lee bounds (Lee, 2009). While Lee bounds are quite conservative estimates and we do not assume that the upper or lower bounds are likely to be observed in the real world, for completeness, we present them here before turning to the question of how likely positive effects might actually be.

As Appendix Section C.5 shows, it is indeed possible that attrition made the effects appear negative in certain specifications, when they might in fact be positive or indistinguishable from 0. When examining just the \$2,000 group, the Lee bound 95% CIs rule out effect sizes higher than 0.09 *SDs* for the financial index, 0.16 *SDs* for the psychological index, 0.23 *SDs* for the cognitive capacity index, and 0.15 *SDs* for the health index (see Appendix Table H.9). Notably, the upper bounds of the 95% CI Lee bounds are still lower than 9 out of 12 (4 indices \times 3 time periods) average expert predictions from our prediction study. This suggests that even making fairly extreme assumptions about missing data yields estimates that are only marginally positive and considerably more pessimistic than most experts believed.

Are positive effects likely? Importantly, however, the bounding exercises also reveal that it is highly unlikely that the true effect of cash could have been meaningfully positive for all the indices. Specifically, using an inverted Horowitz-Manski bounding exercise (Horowitz and Manski, 2000; Baird, McIntosh and Özler, 2019), we calculate how extreme the missing participants' outcomes would have needed to be for us to conclude that cash had positive effects. We find that the gap between the missing cash and missing Control group participants would have needed to be between 0.4 and 0.7 *SDs* (depending on the index)—in the *opposite* direction of what we observe on average—for us to conclude that cash had a positive effect. To put these numbers into context,

depending on the index, the 0.4 to 0.7 *SDs* is equivalent to 1.0–4.7× the (non-causal) “effect” of moving from below pre-treatment median income to above it. See Appendix Section C.5.

What effects are likely based on reasonable imputation? Per our pre-analysis plan, we employ a widely-used multiple imputation approach designed for time-series cross-sectional data (Honaker and King, 2010) to identify an outcome we would be likely to see if we were not missing any data. As Appendix Section C.6 shows, these analyses continue to reveal (weakly) negative effects for all four indices. Given that this approach suggests broadly negative results, and given the various robustness checks above that provide no compelling evidence or rationale to suggest positive results, overall we remain confident in the general conclusion that cash did not have positive effects.

5 Mechanisms

Why did (more) cash not have an observable positive effect on the financial, psychological, cognitive capacity, and health survey outcomes, as predicted by experts and laypeople? Below, we explore seven possible mechanisms. We find relatively little evidence for six of them (Section 5.1). We then focus on the mechanism that appears most consistent with our data—saliency of financial obligations (Section 5.2)—and present a model that formalizes it (Section 5.3).

We note two caveats. First, the analyses below and in Appendix Section D are not preregistered and thus should be interpreted with reasonable caution. Second, while we identify and focus on one mechanism that is most consistent with our data, it is possible that other mechanisms may still play a role. We discuss the (mostly) ruled out mechanisms in part to lay the groundwork for future research.

5.1 Mechanisms (mostly) ruled out

Strategic responding. Although cash recipients were told that they had been randomly chosen to receive the UCTs, they may have nevertheless believed that the non-profit gave money to people who they believed were in greatest need. If the cash groups also believed that they had more to gain financially from making themselves sound needier (relative to the Control group), they (but not the Control group) may have strategically distorted their responses “downward” (Moore, Stinson and Welniak, 2000; Martinelli and Parker, 2009; Baird and Özler, 2012; Beegle et al., 2012). To test this possibility, we embedded an experiment ($N = 2,423$) into the survey, collected correlational survey evidence, and compared self-reports to administrative bank data to test for underreporting of assets. The results from the first two approaches are largely inconsistent with this mechanism, while the results from the final approach are consistent; see Appendix Section D.1.

Reference dependence. Another possibility relates to reference dependence (Kahneman and Tversky, 1979): If the cash groups spent or earmarked their money before taking the post-treatment surveys and compared themselves to a time in which they still had money, they may have felt poorer

than if they had never received cash. We test this hypothesis using a second embedded experiment ($N = 2,474$) and correlational survey responses; see Appendix Section D.2. Overall, the data suggest that there is some, but not particularly strong, evidence for this mechanism.

Spending behavior. We also consider the possibility that UCT recipients spent money in ways that ultimately harmed them, for instance on recreational drugs and alcohol (Banerjee and Du-flo, 2007; Brune et al., 2017). Consistent with prior work (Evans and Popova, 2017; Al Izzati, Suryadarma and Suryahadi, 2023), we find no evidence of increased spending in these categories; see Appendix Section D.3.

Expectations of money amounts. Another possibility relates to expectations of how much (more) money one will receive. We consider three variants of this story. In the first, it could be that the cash groups were disappointed to “only” receive \$500 or \$2,000, which in turn pushed their post-treatment survey responses downward. A second variation is that cash group participants believed that the UCTs would be repeated rather than a one-time windfall and subsequently were disappointed to learn that they were wrong, and/or were disappointed that they were experiencing a declining sequence of payments from \$500 or \$2,000 to only the survey payments (Loewenstein and Prelec, 1993). Finally, it is possible that the Control group expected to receive money throughout the trial, and this hope elevated their survey responses. Overall, as detailed in Appendix Section D.4, we conclude that it is unlikely that such expectations could explain our results.

Inference from money receipt. Perhaps simply receiving money from a charity led UCT recipients to infer that they must be poor, which in turn pushed their survey responses “downward” relative to Control group participants who would not have made this inference. We believe this mechanism is unlikely to have played a large role in our results; see Appendix Section D.5.

Declining social relationships. Another possibility is that receiving cash harmed the participants’ relationships with their friends or family, for instance due to the participants not sharing enough of their funds or the participants distancing themselves from friends and family to avoid being asked for help (Portes, 1998; Dana, Weber and Kuang, 2007; O’Brien, 2012; McNeill and Pierotti, 2021). We find no evidence for this pathway; see Appendix Section D.6.

5.2 Saliency of financial obligations

The final mechanism we explore relates to the saliency of financial obligations. If receiving cash made participants think about the ways in which they could spend that cash—i.e., if it led them to think about existing financial obligations and potentially uncover new ones—they could have been distressed by this, particularly if they found that these obligations were larger than expected and the cash windfall was insufficient to address them. We find several pieces of evidence consistent with this explanation.

Money “on the mind.” We find that the cash groups thought about money more in hypothetical scenarios that were plausibly, but not necessarily, related to money (Shah et al., 2018). In the original work on this topic, Shah et al. found that income was negatively associated with the extent to which people thought about money in these scenarios: the poorer people were, the more they thought about money. They further found that poorer people did not think about non-monetary issues more, providing evidence consistent with the idea that financial issues, but not non-financial issues, captured the attention of people struggling with their finances (Mani et al., 2013; Mullainathan and Shafir, 2013). Based on this work, we expected that providing poor individuals with a positive shock to their finances through a UCT would decrease the extent to which they thought about money. However, we in fact find the opposite: both the \$500 and \$2,000 groups thought about money more rather than less (collapsing across all the scenarios in the post-treatment surveys: $\beta_{\$500} = 0.152, p < 0.001$; $\beta_{\$2000} = 0.144, p = 0.001$).⁴⁰

Needs over the next 30 days. Second, in t2 we asked participants to indicate whether they had enough money to pay for everything their household needed to pay for over the next 30 days and, if not, how much more money they needed. We find that the cash groups indicated that they would need substantially more money than the Control group (Control mean=\$828, $\beta_{\$500} = \$120, p < 0.001$, $\beta_{\$2000} = \$192, p < 0.001$). These results suggest that, relative to the Control group, the cash groups believed they had greater needs or obligations.

Hypothetical stimulus check spending. A third piece of evidence comes from a different t2 question: “Imagine that the government decided to give everyone a \$500 stimulus check. If you got this money today, what are the MAIN thing(s) you would spend the money on?” Participants were then shown 18 categories (e.g., rent, groceries, paying off debts), of which they could select one or more. We find that the cash groups chose significantly more spending categories than Control: $M_{Control} = 2.8, M_{\$500} = 3.0, M_{\$2000} = 3.4$ (both $p \leq 0.009$; see Appendix Figure G.10). One possible interpretation (though not the only one) is that the cash groups had a larger number of financial obligations salient to them.

Overwhelmed by others’ needs. Fourth, participants were asked the extent to which they agreed with the statement, “Over the past week, I have felt overwhelmed or burdened by the financial needs of people outside my household.” We find that relative to Control, the cash groups reported feeling more overwhelmed or burdened by others’ needs post-treatment ($\beta_{\$500} = 0.101, p = 0.014$; $\beta_{\$2000} = 0.214, p < 0.001$). The effect sizes are similar when restricting to the t2 survey ($\beta_{\$500} = 0.129, p = 0.038$; $\beta_{\$2000} = 0.205, p = 0.008$). One interpretation of these results is that cash participants simultaneously viewed supporting friends and family outside the household as a potential financial responsibility or domain in which they would like to spend their money, but

⁴⁰Participants responded to two scenarios in each time period, one developed by Shah et al. (2018) and one developed by us. The results are robust to analyzing only the former rather than both.

perhaps did not feel fully financially able to do so.⁴¹

Spending decision stress. Finally, participants who had received UCTs were more likely to agree with the statement, “Over the past week, I have felt stressed by needing to decide how to spend the money I have” (for all post-treatment surveys: $\beta_{\$500} = 0.201, p < 0.001$; $\beta_{\$2000} = 0.163, p = 0.001$). These results are consistent with the notion that UCT recipients were thinking about their finances and how to optimally allocate their cash windfall. This finding is also consistent with prior work documenting negative effects of choice, e.g., through choice overload (Iyengar and Lepper, 2000) and regret (Sugden, 1985).

Mediation analysis. The five aforementioned variables (money on the mind, needs over the next 30 days, stimulus check spending, overwhelmed by others’ needs, and spending decision stress) either partially or fully statistically mediate the effect of the treatment on the indices, depending on the analysis. Receiving a UCT significantly increased the values for all five variables (all $p \leq 0.014$), which in turn have either a negative or no relationship with the four indices when controlling for treatment group (for 15 out of 19 coefficients: $\beta < 0, p \leq 0.021$; for 4 out of 19: NS ; “money on the mind” is not included as a mediator for the cognitive capacity index because of collinearity with the index). Finally, when we add the mediators to the regressions of index on UCT indicators, the effects of the UCTs on the indices weaken and sometimes lose significance. Appendix Table H.11 shows the mediation using the three variables that were measured in all post-treatment time periods. These analyses suggest that the saliency of financial obligations, which may have stemmed from a deeper engagement with one’s finances, could have played a role in explaining the non-positive treatment effects; however, we urge caution in viewing these as necessarily causal paths given the inherent limitations of mediation analyses (Fiedler, Schott and Meiser, 2011; Celli, 2022).

5.3 Model

To better elucidate the mechanism that is most consistent with our data and to explain how and when it could lead to lower well-being for some people (at least in the short run), we propose a relatively simple discrete time model that captures how an agent allocates scarce resources to manage their finances. The basic structure of the model focuses on the fact that (re-)optimizing one’s financial decisions in the face of new information and financial shocks has benefits, but also incurs costs, and hence that there is a tradeoff between “passively” following a predetermined plan and “actively” engaging with a potentially altered financial portfolio.

In the context of our field experiment, the model formalizes how receiving a positive financial windfall—referred to as a “bonus” below—can lead to an improved reallocation of resources new and old, but also to a realization that obligations that were previously (rationally) neglected are more serious than anticipated. As a result, agents in our model may initially experience net negative

⁴¹These responses could also reflect social conventions or social desirability bias.

utility after receiving a windfall payment, despite the fact that the money has a positive direct effect on their cashflow and debt repayment.

In addition to capturing key features and findings from our empirical study, the model also makes predictions that go beyond our current experiment and can help inform future studies. Below, we describe the model setup, the decision problem the agent is facing, and the intuition that we can derive from this model; the details are in Appendix Section I.

Setup. We study the financial management strategy of an agent who can choose to take a passive or active approach towards the repayment of a stock of debt and obligations, denoted by D . Taking an active approach towards some debt and obligations involves paying an associated cost a to be able to observe any changes that might occur involving these debts and obligations. We assume a three-period setting, where the agent earns income 0, M_1 , and M_2 in periods 0, 1, and 2, respectively; all variables are real and non-negative.

In period 0, which can be thought of purely as a planning phase, no money is earned and no payments are made. The agent provisionally assumes that they will be passive in period 1—it is generally optimal for the agent not to deal with their finances every period but rather to optimize an initial plan and then carry it out. In period 0, they decide what payment they would like to make towards the debt, denoted by \bar{d} . Unpaid debt or obligations may accrue additional costs over time, which—when dealt with passively—occur without the agent observing these accumulating costs (we will capture the idea of these negative financial developments more generally through a “shock” below). These accumulating costs could come, for instance, in the form of late fees, interest payments, or small problems becoming more serious over time (due to, e.g., deferring maintenance on car repairs or delaying preventative healthcare).

Period 1 stands in for a typical financial cycle (e.g., a month). In the background, Nature introduces a (negative) shock S at the beginning of period 1, which can take a value of either 0 or $s > 0$, with probabilities $1 - q$ and q , respectively. S captures negative events such as individual economic conditions worsening (e.g., late fees), but also general economic downturns (e.g., interest rates rising). Independently and more rarely, Nature also introduces a possible monetary bonus subsequently in period 1, which can take a value of either 0 or $b > 0$, with probabilities $1 - p$ and p respectively. The bonus represents a positive windfall the agent might experience, such as the (probably somewhat rare) unconditional cash transfer from our treatment. The agent always observes the realization of B , but not necessarily the existence or true magnitude of the shock S . The associated probabilities p and q , respectively, are known to the agent.

Decision problem. After observing B at the start of period 1, the agent must choose whether to continue with this passive strategy, in which case the initial debt payment \bar{d} will be implemented and prior expectations about the size of the shock remain in place, or to switch to an active strategy. If the agent chooses the active strategy, there is an associated fixed cost a , which may reflect economic costs such as time and effort, but may also include psychological costs such as facing potentially aversive information (Karlsson, Loewenstein and Seppi, 2009; Golman, Hagmann and Loewenstein,

2017; Andries and Haddad, 2020). In choosing the active strategy, they also learn the true realized value of S and subsequently have the opportunity to re-optimize the payment towards the total debt, now choosing d^* .

The agent derives utility from consumption, with a per-period utility function $u(\cdot)$, which is assumed to be concave. As detailed in Appendix Section I, for tractability the utility function is assumed to be isoelastic with parameter η ; however, any concave function will lead to similar dynamics. The agent's objective is to maximize their expected utility, and hence to equalize consumption across periods, *ceteris paribus*. Their problem is characterized by the initial choice of \bar{d} in period 0, the decision to take an active or passive stance in debt management in period 1, and the choice of d^* if they decided to be active, also in period 1. Period 2 occurs much later and, during this final period, all agents are forced to pay a and any remaining debt. Essentially, period 2 is the point of reckoning, when all payments come in and go out and must balance.

Assumptions. Our model uses as a starting point a fully rational decision-maker who optimizes their expected utility over their lifetime. We keep our model classical in almost all regards with only one exception: we introduce the possibility of a behavioral type of agent that differs from the “benchmark” type only insofar as they mispredict the magnitude of the shock. While the benchmark agent correctly believes $\tilde{s} = s$, the behavioral type optimistically believes $\tilde{s} < s$. Our model makes minimal structural or behavioral assumptions about this systematic deviation from s . We introduce this behavioral type building on a wide range of behavioral patterns documented in the financial attention and decision-making literature. Behavioral patterns that could explain why an agent might have $\tilde{s} < s$ include overoptimism of avoiding a negative shock (Brunnermeier and Parker, 2005; Howard et al., 2022) and mispredicting the ability to repay (growing) debt and obligations (Stango and Zinman (2009); Leary and Wang (2016); although see Allcott et al. (2022)).

Numerical approach. The model posits three periods with multiple decision points over time that may be contingent on the (potentially unobserved) realizations of B and S . We solve this decision problem with a numerical approach, using backward induction. To do so, we fix the value of several parameters in the model (across a spectrum of feasible levels) and only vary one key parameter at a time, including a (the cost the agent has to pay to actively manage their otherwise passive obligations) and \tilde{s} .

We focus on the natural range of parameter values for which it is in the agent's interest to actively manage more of their obligations only after receiving a windfall $B = b$. (By construction, $a > 0$ implies that agents will not choose an active debt management strategy if $B = 0$ since they can already plan a maximizing value of \bar{d} for precisely that scenario.) The exact parameter values, and a demonstration of robustness to a wider range of values, are in Appendix Section I.

Solution. In solving the model, we focus on the utility experienced by agents from the perspective of period 1—after receiving, or not receiving, the cash windfall—as that is the most pertinent comparison to our field experiment results. Note that our model also enables us to speak to later

experienced utility (in period 2), which is outside the scope of our empirical setting but may prove useful for future research.

The findings from the model can be summarized as follows. For many values of a , both the benchmark and behavioral types choose to actively manage their debt when they receive the windfall, but not otherwise. Following their observation of the realized S , any agent who chose to be active re-optimizes their debt payments to d^* . However, while the benchmark type expected to find S to be the true magnitude s , the behavioral type expected $\tilde{s} < s$. For a nontrivial range of reasonable parameters, this unhappy surprise more than offsets the positive impact of receiving $B = b$, causing this agent to have lower utility in period 1 than their counterpart who did not get a windfall (and hence also continues to expect a smaller shock).⁴² Appendix Section I, and Appendix Figures I.14 and I.15, shows the robustness of these results for a wide range of parameter values.

Interpretation and further predictions. Our model rationalizes and offers one possible explanation for why participants in our trial who received a UCT might have reported worse outcomes than participants who did not get a UCT. The model suggests that (behavioral) agents can experience negative utility shortly after receiving the cash windfall because, by choosing to re-optimize their debt management strategy after the windfall, they learn that they have more obligations than they previously thought, and thus they cannot consume as much as planned. Of course, this may not be true of everyone in our sample, but it can rationalize our aggregated empirical findings and shed new light on the potential challenges arising for low-income individuals when receiving cash transfers.

Importantly, our model also makes two predictions that we are not able to test in our current empirical setting, but that offer guidance for future research. First, the model predicts that sufficiently high cash payments will *not* result in negative utility in period 1. This is because, once the cash transfer is large enough, it will be sufficient to pay all obligations and as a result lead to positive utility in period 1 from having received the cash. Second, even if the cash is insufficient to cover all obligations in period 1, our model would still suggest that lifetime utility is optimized through the cash transfer. This means that agents should be better off (or, at the very least, not worse off) in the long run because they settled some of their obligations earlier, and thus those obligations did not get worse over time. The long-term view is represented by period 2 in our model, which captures some distant point in the future when all obligations are due, but not a time period we have available in our data. However, a future study may wish to measure and explore this prediction empirically.

6 Discussion

This paper reported on a randomized controlled trial that provided people experiencing poverty with nothing, \$500, or \$2,000. The data reveal that participants spent the cash windfall fairly quickly,

⁴²The benchmark type, on the other hand, is accurate in their perception of S and therefore does not experience an unhappy surprise after paying a ; consequently, their utility is not negatively affected in period 1.

with increased expenditures dwindling down the UCTs within a matter of weeks. Surprisingly, this increased ability to spend money did not translate to positive differences in survey outcomes at any time point — even only one week after cash receipt. In particular, we found no positive differences in financial, psychological, cognitive capacity, or health survey outcomes, neither between the Control and cash groups, nor between the two cash groups. If anything, the results show that, relative to the Control group, cash groups reported worse financial, psychological, and health (but not cognitive capacity) outcomes, particularly for more subjectively-measured outcomes. Once we account for attrition, we cannot generally reject a null effect of cash on our outcomes but, at the same time, the analyses also reveal that it is highly unlikely that the effects could have been positive. In seeking to explain these non-positive results, we tested a number of potential mechanisms. Although several explanations may be possible, our data is most consistent with the notion—further elaborated through our model—that the windfall led participants to engage with their finances and think about their (extensive) financial obligations, which in turn generated distress.

The non-positive effects of the UCTs on survey outcomes stand in stark contrast to the incentivized predictions of both experts and laypeople, who on average dramatically overestimated how effective the UCTs would be at improving outcomes. Overall, our results suggest that our posteriors about the effects of similar future cash transfer policies in similar settings should be somewhat more muted than many—including experts—would have previously believed.

Our findings should be interpreted in light of several caveats. While we aimed to be relatively comprehensive in our outcome measurements, we may have missed some important positive effects of cash, such as time investments into human capital or children’s development. Relatedly, there may have been some positive elements to outcomes we treated as negative. For instance, as our model implies, subjective feelings of financial stress may in some cases still be net positive if they encourage a person to address financial problems before they balloon. Finally, our study was conducted during a particularly challenging economic and public health crisis (the COVID-19 pandemic)—a feature that is shared by several contemporaneous studies on this topic. Although this should not affect the internal validity of the study, it may affect the extent to which some of our results are generalizable to other contexts. Nevertheless, we believe our findings raise important questions—and begin to provide some answers—on poverty alleviation and the basic functioning of cash transfers in high-income countries.

First, on which outcome(s) should poverty alleviation programs be optimizing? In particular, what should be the relative importance of objective financial outcomes (e.g., the ability to pay for pressing needs, pay down debt, or save) versus subjective well-being (e.g., how anxious or stressed a person feels)? If the goal is to increase the former, then simply providing cash to those in need likely accomplishes that goal. If the goal is, at least in part, to increase the latter, then the results from this study suggest that unrestricted one-off UCT payments of this magnitude in such settings may not always be the correct tool.

What, then, might the correct tool be? Budget constraints notwithstanding, one option might be to increase the amount of money given. Indeed, our model predicts that the observed negative

psychological effects should disappear for sufficiently large cash transfers. What constitutes “sufficiently large”? Consider that the average cash transfer amount in the Global South is 37% of the country’s median per capita annual income (Dwyer, Stewart and Zhao, 2023). In contrast, participants in the \$2,000 condition in our study received the equivalent of about 5% of the country’s median per capita annual income.⁴³ Given that participant income and savings were so low, one could argue that even 5% should be helpful. However, it seems similarly plausible that precisely because participant income was so low, needs were also vast, and thus the UCT amounts could have been swamped by those needs. Consistent with this, after the study concluded, we conducted qualitative interviews with 15 participants, where, among other things, we asked them how much money they would have needed to receive for it to have made a difference in their lives at that time. About half of participants indicated that they would have needed more than \$2,000, with those responses falling between \$4,000 and \$25,000. While these responses are highly anecdotal, they at least offer some indication of the perceived magnitude of the need, and underscore the importance of considering the size of cash transfers.

Naturally, however, dramatically increasing cash transfer amounts would be very expensive. A different approach to supporting low-income households might therefore be to couple cash transfers with other (potentially more cost-effective and/or complementary) resources, such as investments at the community level or mental health support. This would be consistent with the “cash-plus” approach, which argues that combining cash with other resources and services can be more effective than providing cash alone (Blattman, Jamison and Sheridan, 2017; Sedlmayr, Shah and Sulaiman, 2020; Little et al., 2021; Banerjee et al., 2022). Furthermore, the way in which cash is delivered (e.g., payment timing) could be varied to better match the needs and preferences of the recipients (Kansikas, Mani and Niehaus, 2023). Alternatively, one could support low-income households primarily through other means like in-kind benefits, opportunities for rewarding work (Hussam et al., 2022), and/or more structural change. Future work conducting cost-benefit analyses on these alternatives could help provide a clearer picture.

It is important to note that even if insufficiently large cash windfalls produce no positive effects on subjective well-being, the benefits of UCT programs may still outweigh the costs. This may be particularly true if one considers positive externalities on others, such as the recipients’ children or friends, and/or the recipients’ preferences (Bursztyn, 2016; Liscow and Pershing, 2022). It is our very strong suspicion is that if a group of low-income people (or, for that matter, any group of people) were given the option to have \$0, \$500, or \$2,000, nearly all would choose the \$2,000—even if they knew that it could have no or negative effects on subjective outcomes.

⁴³US Census, using 2021 dollars: <https://www.census.gov/quickfacts/fact/table/US/SEX255219>; accessed 6 October 2023.

References

- Ackerman, Norleen, and Beatrice Paolucci.** 1983. “Objective and Subjective Income Adequacy: Their Relationship to Perceived Life Quality Measures.” *Social Indicators Research*, 12(1): 25–48.
- Aggarwal, Shilpa, Jenny C. Aker, Dahyeon Jeong, Naresh Kumar, David Sungho Park, Jonathan Robinson, and Alan Spearot.** 2024. “The Dynamic Effects of Cash Transfers to Agricultural Households.” *National Bureau of Economic Research*, w32431.
- Aiken, Emily, Suzanne Bellue, Joshua Blumenstock, Dean Karlan, and Christopher R. Udry.** 2023. “Estimating Impact with Surveys versus Digital Traces: Evidence from Randomized Cash Transfers in Togo.” *National Bureau of Economic Research*, w31751.
- Aizer, Anna, Sungwoo Cho, Shari Eli, and Adriana Lleras-Muney.** 2024. “The Impact of Cash Transfers to Poor Mothers on Family Structure and Maternal Well-Being.” *American Economic Journal: Applied Economics*, 16(2): 492–529.
- Akee, Randall K. Q., William E. Copeland, Gordon Keeler, Adrian Angold, and E. Jane Costello.** 2010. “Parents’ Incomes and Children’s Outcomes: A Quasi-Experiment Using Transfer Payments from Casino Profits.” *American Economic Journal: Applied Economics*, 2(1): 86–115.
- Al Izzati, Ridho, Daniel Suryadarma, and Asep Suryahadi.** 2023. “Do Short-Term Unconditional Cash Transfers Change Behaviour and Preferences? Evidence from Indonesia.” *Oxford Development Studies*, 51(3): 291–306.
- Allcott, Hunt, Joshua Kim, Dmitry Taubinsky, and Jonathan Zinman.** 2022. “Are High-Interest Loans Predatory? Theory and Evidence from Payday Lending.” *The Review of Economic Studies*, 89(3): 1041–1084.
- Andersen, Asbjørn G., Andreas Kotsadam, and Vincent Somville.** 2022. “Material Resources and Well-Being — Evidence from an Ethiopian Housing Lottery.” *Journal of Health Economics*, 83: 102619.
- Anderson, Michael L.** 2008. “Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects.” *Journal of the American Statistical Association*, 103(484): 1481–1495.
- Andries, Marianne, and Valentin Haddad.** 2020. “Information Aversion.” *Journal of Political Economy*, 128(5): 1901–1939.
- Baird, Sarah, and Berk Özler.** 2012. “Examining the Reliability of Self-Reported Data on School Participation.” *Journal of Development Economics*, 98(1): 89–93.
- Baird, Sarah, Craig McIntosh, and Berk Özler.** 2011. “Cash or Condition? Evidence from a Cash Transfer Experiment.” *The Quarterly Journal of Economics*, 126(4): 1709–1753.
- Baird, Sarah, Craig McIntosh, and Berk Özler.** 2019. “When the Money Runs out: Do Cash Transfers Have Sustained Effects on Human Capital Accumulation?” *Journal of Development Economics*, 140: 169–185.
- Baird, Sarah, David McKenzie, and Berk Özler.** 2018. “The Effects of Cash Transfers on Adult Labor Market Outcomes.” *IZA Journal of Development and Migration*, 8(1): 22.
- Baker, Scott, Robert A Farrokhnia, Steffen Meyer, Michaela Pagel, and Constantine Yannelis.** 2023. “Income, Liquidity, and the Consumption Response to the 2020 Economic Stimulus Payments.” *Review of Finance*, 27(6): 2271–2304.
- Banerjee, Abhijit, Dean Karlan, Robert Osei, Hannah Trachtman, and Christopher Udry.** 2022. “Unpacking a Multi-Faceted Program to Build Sustainable Income for the Very Poor.” *Journal of Development Economics*, 155: 102781.
- Banerjee, Abhijit, Michael Faye, Alan Krueger, Paul Niehaus, and Tavneet Suri.** 2023.

- “Universal Basic Income: Short-Term Results from a Long-Term Experiment in Kenya.”
- Banerjee, Abhijit V, and Esther Duflo.** 2007. “The Economic Lives of the Poor.” *Journal of Economic Perspectives*, 21(1): 141–167.
- Bartos, František, Maximilian Maier, T. D. Stanley, and Eric-Jan Wagenmakers.** 2022. “Adjusting for Publication Bias Reveals Mixed Evidence for the Impact of Cash Transfers on Subjective Well-Being and Mental Health.” *PsyArXiv*.
- Beegle, Kathleen, Joachim De Weerdt, Jed Friedman, and John Gibson.** 2012. “Methods of Household Consumption Measurement through Surveys: Experimental Results from Tanzania.” *Journal of Development Economics*, 98(1): 3–18.
- Benjamini, Yoav, and Daniel Yekutieli.** 2001. “The Control of the False Discovery Rate in Multiple Testing under Dependency.” *The Annals of Statistics*, 29(4): 1165–1188.
- Benjamini, Yoav, and Yosef Hochberg.** 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1): 289–300.
- Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, and Bertil Tungodden.** 2015. “Human and Financial Capital for Microenterprise Development: Evidence from a Field and Lab Experiment.” *Management Science*, 61(4): 707–722.
- Bilker, Warren B., John A. Hansen, Colleen M. Bresniger, et al.** 2012. “Development of Abbreviated Nine-Item Forms of the Raven’s Standard Progressive Matrices Test.” *Assessment*, 19(3): 354–369.
- Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan.** 2017. “Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia.” *American Economic Review*, 107(4): 1165–1206.
- Blattman, Christopher, Nathan Fiala, and Sebastian Martinez.** 2014. “Generating Skilled Self-Employment in Developing Countries: Experimental Evidence from Uganda.” *The Quarterly Journal of Economics*, 129(2): 697–752.
- Braveman, Paula A., Catherine Cubbin, Susan Egerter, David R. Williams, and Elsie Pamuk.** 2010. “Socioeconomic Disparities in Health in the United States: What the Patterns Tell Us.” *American Journal of Public Health*, 100(S1): S186–S196.
- Brooks, Wyatt, Kevin Donovan, Terence R Johnson, and Jackline Oluoch-Aridi.** 2022. “Cash Transfers as a Response to COVID-19: Experimental Evidence from Kenya.” *Journal of Development Economics*, 158: 29.
- Brune, Lasse, Xavier Giné, Jessica Goldberg, and Dean Yang.** 2017. “Savings Defaults and Payment Delays for Cash Transfers: Field Experimental Evidence from Malawi.” *Journal of Development Economics*, 129: 1–13.
- Brunnermeier, Markus K., and Jonathan A. Parker.** 2005. “Optimal Expectations.” *American Economic Review*, 95(4): 1092–1118.
- Bursztyn, Leonardo.** 2016. “Poverty and the Political Economy of Public Education Spending: Evidence from Brazil.” *Journal of the European Economic Association*, 14(5): 1101–1128.
- Camerer, Colin F., Anna Dreber, Felix Holzmeister, et al.** 2018. “Evaluating the Repliability of Social Science Experiments in Nature and Science between 2010 and 2015.” *Nature Human Behaviour*, 2(9): 637–644.
- Cañedo, Ana P., Raissa Fabregas, and Prankur Gupta.** 2023. “Emergency Cash Transfers for Informal Workers: Impact Evidence from Mexico.” *Journal of Public Economics*, 219: 104820.
- Carvalho, Leandro, Arna Olafsson, and Dan Silverman.** 2024. “Misfortune and Mistake: The Financial Conditions and Decision-making Ability of High-Cost Loan Borrowers.” *Journal of Political Economy*, 730200.
- Carvalho, Leandro S., Stephan Meier, and Stephanie W. Wang.** 2016. “Poverty and Eco-

- nomic Decision-Making: Evidence from Changes in Financial Resources at Payday.” *American Economic Review*, 106(2): 260–284.
- Celli, Viviana.** 2022. “Causal Mediation Analysis in Economics: Objectives, Assumptions, Models.” *Journal of Economic Surveys*, 36(1): 214–234.
- Cesarini, David, Erik Lindqvist, Robert Östling, and Björn Wallace.** 2016. “Wealth, Health, and Child Development: Evidence from Administrative Data on Swedish Lottery Players.” *The Quarterly Journal of Economics*, 131(2): 687–738.
- CFPB.** 2017a. “CFPB Financial Well-Being Scale: Scale Development Technical Report.” Consumer Financial Protection Bureau.
- CFPB.** 2017b. “National Financial Well-Being Survey: Public Use File Codebook.”
- Chetty, Raj, John N Friedman, Michael Stepner, and Opportunity Insights Team.** 2024. “The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data.” *The Quarterly Journal of Economics*, 139(2): 829–889.
- Christian, Cornelius, Lukas Hensel, and Christopher Roth.** 2019. “Income Shocks and Suicides: Causal Evidence From Indonesia.” *The Review of Economics and Statistics*, 101(5): 905–920.
- Covarrubias, Rebecca, Andrea Romero, and Michael Trivelli.** 2015. “Family Achievement Guilt and Mental Well-being of College Students.” *Journal of Child and Family Studies*, 24(7): 2031–2037.
- Curl, Heather, Annette Lareau, and Tina Wu.** 2018. “Cultural Conflict: The Implications of Changing Dispositions Among the Upwardly Mobile.” *Sociological Forum*, 33(4): 877–899.
- Dahl, Gordon B., and Lance Lochner.** 2012. “The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit.” *American Economic Review*, 102(5): 1927–1956.
- Dana, Jason, Roberto A. Weber, and Jason Xi Kuang.** 2007. “Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness.” *Economic Theory*, 33(1): 67–80.
- de Bruijn, Ernst-Jan, and Gerrit Antonides.** 2022. “Poverty and Economic Decision Making: A Review of Scarcity Theory.” *Theory and Decision*, 92(1): 5–37.
- DellaVigna, Stefano, Devin Pope, and Eva Vivalt.** 2019. “Predict Science to Improve Science.” *Science*, 366(6464): 428–429.
- Destin, Mesmin, and Régine Debrosse.** 2017. “Upward Social Mobility and Identity.” *Current Opinion in Psychology*, 18: 99–104.
- Dreber, Anna, Thomas Pfeiffer, Johan Almenberg, et al.** 2015. “Using Prediction Markets to Estimate the Reproducibility of Scientific Research.” *Proceedings of the National Academy of Sciences*, 112(50): 15343–15347.
- Dwyer, Ryan, Anita Palepu, Claire Williams, Daniel Daly-Grafstein, and Jiaying Zhao.** 2023. “Unconditional Cash Transfers Reduce Homelessness.” *Proceedings of the National Academy of Sciences*, 120(36): e2222103120.
- Dwyer, Ryan J., and Elizabeth W. Dunn.** 2022. “Wealth Redistribution Promotes Happiness.” *Proceedings of the National Academy of Sciences*, 119(46): e2211123119.
- Dwyer, Ryan, Kaitlyn Stewart, and Jiaying Zhao.** 2023. “A Comparison of Cash Transfer Programs in the Global North and South.” In *Cash Transfers for Inclusive Societies: A Behavioral Lens*. University of Toronto Press.
- Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker.** 2022. “General Equilibrium Effects of Cash Transfers: Experimental Evidence From Kenya.” *Econometrica*, 90(6): 2603–2643.
- Erten, Bilge, Pinar Keskin, and Silvia Prina.** 2022. “Social Distancing, Stimulus Payments,

- and Domestic Violence: Evidence from the US during COVID-19.” *AEA Papers and Proceedings*, 112: 262–266.
- Evans, David K., and Anna Popova.** 2017. “Cash Transfers and Temptation Goods.” *Economic Development and Cultural Change*, 65(2): 189–221.
- Evans, William N., and Timothy J. Moore.** 2011. “The Short-Term Mortality Consequences of Income Receipt.” *Journal of Public Economics*, 95(11-12): 1410–1424.
- Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde.** 2022. “The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences.” *Management Science*.
- Fiedler, Klaus, Malte Schott, and Thorsten Meiser.** 2011. “What Mediation Analysis Can (Not) Do.” *Journal of Experimental Social Psychology*, 47(6): 1231–1236.
- Field, Erica M., and Elisa M. Maffioli.** 2021. “Are Behavioral Change Interventions Needed to Make Cash Transfer Programs Work for Children? Experimental Evidence from Myanmar.” *National Bureau of Economic Research*, w28443.
- FragileFamilies.** 2011. “Fragile Families Survey Instrument - Moms Yr 9.”
- Friedman, Sam.** 2016. “Habitus Clivé and the Emotional Imprint of Social Mobility.” *The Sociological Review*, 64(1): 129–147.
- Gallup.** 2017. “Gallup Daily Methodology.”
- Gardner, Jonathan, and Andrew J. Oswald.** 2007. “Money and Mental Wellbeing: A Longitudinal Study of Medium-Sized Lottery Wins.” *Journal of Health Economics*, 26(1): 49–60.
- Gasiorowska, Agata.** 2014. “The Relationship between Objective and Subjective Wealth Is Moderated by Financial Control and Mediated by Money Anxiety.” *Journal of Economic Psychology*, 43: 64–74.
- Gennetian, Lisa A., Greg J. Duncan, Nathan A. Fox, Sarah Halpern-Meekin, Katherine Magnuson, Kimberly G. Noble, and Hirokazu Yoshikawa.** 2024. “Effects of a Monthly Unconditional Cash Transfer Starting at Birth on Family Investments among US Families with Low Income.” *Nature Human Behaviour*, 1–16.
- Ghanem, Dalia, Sarojini Hirshleifer, and Karen Ortiz-Beccera.** 2023. “Testing Attrition Bias in Field Experiments.” *Journal of Human Resources*.
- Giurge, Laura M., Ashley V. Whillans, and Colin West.** 2020. “Why Time Poverty Matters for Individuals, Organisations and Nations.” *Nature Human Behaviour*, 4(10): 993–1003.
- Gladstone, Joe J., Jon M. Jachimowicz, Adam Eric Greenberg, and Adam D. Galinsky.** 2021. “Financial Shame Spirals: How Shame Intensifies Financial Hardship.” *Organizational Behavior and Human Decision Processes*, 167: 42–56.
- Godoy, Ricardo, Dean Karlan, and Jonathan Zinman.** 2021. “Randomization for Causality, Ethnography for Mechanisms: Illiquid Savings for Liquor in an Autarkic Society.” *National Bureau of Economic Research*, w29566.
- Golman, Russell, David Hagmann, and George Loewenstein.** 2017. “Information Avoidance.” *Journal of Economic Literature*, 55(1): 96–135.
- Golosov, Mikhail, Michael Graber, Magne Mogstad, and David Novgorodsky.** 2024. “How Americans Respond to Idiosyncratic and Exogenous Changes in Household Wealth and Unearned Income*.” *The Quarterly Journal of Economics*, 139(2): 1321–1395.
- Gupta, Prankur, Daniel Stein, Kyla Longman, Heather Lanthorn, Rico Bergmann, Emmanuel Nshakira-Rukundo, Noel Rutto, Christine Kahura, Winfred Kananu, Gabrielle Posner, K.J. Zhao, and Penny Davis.** 2024. “Cash Transfers amid Shocks: A Large, One-Time, Unconditional Cash Transfer to Refugees in Uganda Has Multidimensional Benefits after 19 Months.” *World Development*, 173: 106339.
- Handa, Sudhanshu, Luisa Natali, David Seidenfeld, Gelson Tembo, and Benjamin**

- Davis.** 2018. “Can Unconditional Cash Transfers Raise Long-Term Living Standards? Evidence from Zambia.” *Journal of Development Economics*, 133: 42–65.
- Haushofer, Johannes, and Ernst Fehr.** 2014. “On the Psychology of Poverty.” *Science*, 344(6186): 862–867.
- Haushofer, Johannes, and Jeremy Shapiro.** 2016. “The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya.” *The Quarterly Journal of Economics*, 131(4): 1973–2042.
- Haushofer, Johannes, Charlotte Ringdal, Jeremy P. Shapiro, and Xiao Yu Wang.** 2019. “Income Changes and Intimate Partner Violence: Evidence from Unconditional Cash Transfers in Kenya.” *National Bureau of Economic Research*, w25627.
- Haushofer, Johannes, Robert Mudida, and Jeremy Shapiro.** 2023. “The Comparative Impact of Cash Transfers and a Psychotherapy Program on Psychological and Economic Well-being.”
- Hidrobo, Melissa, Amber Peterman, and Lori Heise.** 2016. “The Effect of Cash, Vouchers, and Food Transfers on Intimate Partner Violence: Evidence from a Randomized Experiment in Northern Ecuador.” *American Economic Journal: Applied Economics*, 8(3): 284–303.
- Honaker, James, and Gary King.** 2010. “What to Do about Missing Values in Time-Series Cross-Section Data.” *American Journal of Political Science*, 54(2): 561–581.
- Horowitz, Joel L., and Charles F. Manski.** 2000. “Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data.” *Journal of the American Statistical Association*, 95(449): 77–84.
- Howard, Ray Charles “Chuck”, David J. Hardisty, Abigail B. Sussman, and Marcel F. Lukas.** 2022. “Understanding and Neutralizing the Expense Prediction Bias: The Role of Accessibility, Typicality, and Skewness.” *Journal of Marketing Research*, 59(2): 435–452.
- Hurst, Allison L.** 2010. *The Burden of Academic Success: Managing Working-Class Identities in College*. Lexington Books.
- Hussam, Reshmaan, Erin M. Kelley, Gregory Lane, and Fatima Zahra.** 2022. “The Psychosocial Value of Employment: Evidence from a Refugee Camp.” *American Economic Review*, 112(11): 3694–3724.
- Iyengar, Sheena S., and Mark R. Lepper.** 2000. “When Choice Is Demotivating: Can One Desire Too Much of a Good Thing?” *Journal of Personality and Social Psychology*, 79(6): 995–1006.
- Jachimowicz, Jon M., Erin L. Frey, Sandra C. Matz, Bertus F. Jeronimus, and Adam D. Galinsky.** 2021. “The Sharp Spikes of Poverty: Financial Scarcity Is Related to Higher Levels of Distress Intensity in Daily Life.” *Social Psychological and Personality Science*.
- Jacob, Brian, Natasha Pilkauskas, Elizabeth Rhodes, Katherine Richard, and H Luke Shaefer.** 2022. “The COVID Cash Transfer Study II: The Hardship and Mental Health Impacts of an Unconditional Cash Transfer to Low-Income Individuals.” *The National Tax Journal*, 75(3): 56.
- Johnson, David S., Jonathan A. Parker, and Nicholas S. Souleles.** 2006. “Household Expenditure and the Income Tax Rebates of 2001.” *American Economic Review*, 96(5): 1589–1610.
- Jones, Damon, David Molitor, and Julian Reif.** 2019. “What Do Workplace Wellness Programs Do? Evidence from the Illinois Workplace Wellness Study.” *The Quarterly Journal of Economics*, 134(4): 1747–1791.
- Kahneman, Daniel, and Amos Tversky.** 1979. “Prospect Theory: An Analysis of Decision under Risk.” *Econometrica*, 47(2): 263–292.
- Kahneman, Daniel, and Angus Deaton.** 2010. “High Income Improves Evaluation of Life but

- Not Emotional Well-Being.” *Proceedings of the National Academy of Sciences*, 107(38): 16489–16493.
- Kansikas, Carolina, Anandi Mani, and Paul Niehaus.** 2023. “Customized Cash Transfers: Financial Lives and Cash-flow Preferences in Rural Kenya.” *National Bureau of Economic Research*, w30930.
- Karger, Ezra, and Aastha Rajan.** 2020. “Heterogeneity in the Marginal Propensity to Consume: Evidence from Covid-19 Stimulus Payments.” *Federal Reserve Bank of Chicago*, Working Paper 2020-15.
- Karlan, Dean, Matt Lowe, Robert Darko Osei, Isaac Osei-Akoto, Benjamin Roth, and Christopher Udry.** 2022. “Social Protection and Social Distancing During the Pandemic: Mobile Money Transfers in Ghana.” *National Bureau of Economic Research*, w30309.
- Karlsson, Niklas, George Loewenstein, and Duane Seppi.** 2009. “The Ostrich Effect: Selective Attention to Information.” *Journal of Risk and Uncertainty*, 38(2): 95–115.
- Kaur, Supreet, Sendhil Mullainathan, Suanna Oh, and Frank Schilbach.** 2021. “Do Financial Concerns Make Workers Less Productive?” *National Bureau of Economic Research*, 28338.
- Kent, Christina, and Alejandro Martínez-Marquina.** 2022. “When a Town Wins the Lottery: Evidence from Spain.” *Working Paper*.
- Kluender, Raymond, Neale Mahoney, Francis Wong, and Wesley Yin.** 2024. “The Effects of Medical Debt Relief: Evidence from Two Randomized Experiments.” *National Bureau of Economic Research*, w32315.
- Kovski, Nicole, Natasha V. Pilkauskas, Katherine Michelmore, and H. Luke Shaefer.** 2023. “Unconditional Cash Transfers and Mental Health Symptoms among Parents with Low Incomes: Evidence from the 2021 Child Tax Credit.” *SSM - Population Health*, 22: 101420.
- Kroenke, Kurt, Robert L. Spitzer, and Janet B. W. Williams.** 2001. “The PHQ-9.” *Journal of General Internal Medicine*, 16(9): 606–613.
- Kuhn, Peter, Peter Kooreman, Adriaan Soeteven, and Arie Kapteyn.** 2011. “The Effects of Lottery Prizes on Winners and Their Neighbors: Evidence from the Dutch Postcode Lottery.” *American Economic Review*, 101(5): 2226–2247.
- Lachman, Margie E, and Suzanne L Weaver.** 1998. “The Sense of Control as a Moderator of Social Class Differences in Health and Well-Being.” *Journal of Personality and Social Psychology*, 74(3): 763–773.
- Leana, Carrie R., and Jirs Meuris.** 2015. “Living to Work and Working to Live: Income as a Driver of Organizational Behavior.” *Academy of Management Annals*, 9(1): 55–95.
- Leary, Jesse B, and Jialan Wang.** 2016. “Liquidity Constraints and Budgeting Mistakes: Evidence from Social Security Recipients.” *Working Paper*.
- Lee, David S.** 2009. “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects.” *The Review of Economic Studies*, 76: 1071–1102.
- Lee, Elizabeth M., and Rory Kramer.** 2013. “Out with the Old, In with the New? Habitus and Social Mobility at Selective Colleges.” *Sociology of Education*, 86(1): 18–35.
- Liebman, Jeffrey, Kathryn Carlson, Eliza Novick, and Pamela Portocarrero.** 2022. “Chelsea Eats Program: Experimental Impacts.” Rappaport Institute for Greater Boston Working Paper.
- Liscow, Zachary, and Abigail Pershing.** 2022. “Why Is So Much Redistribution In-Kind and Not in Cash? Evidence from a Survey Experiment.” *National Tax Journal*, 75(2): 313–354.
- List, John A., Sally Sadoff, and Mathis Wagner.** 2011. “So You Want to Run an Experiment, Now What? Some Simple Rules of Thumb for Optimal Experimental Design.” *Experimental Economics*, 14(4): 439–457.

- Little, Madison T., Keetie Roelen, Brittany C. L. Lange, et al.** 2021. "Effectiveness of Cash-plus Programmes on Early Childhood Outcomes Compared to Cash Transfers Alone: A Systematic Review and Meta-Analysis in Low- and Middle-Income Countries." *PLOS Medicine*, 18(9): e1003698.
- Loewenstein, George F., and Dražen Prelec.** 1993. "Preferences for Sequences of Outcomes." *Psychological Review*, 100: 91–108.
- Londoño-Vélez, Juliana, and Pablo Querubín.** 2022. "The Impact of Emergency Cash Assistance in a Pandemic: Experimental Evidence from Colombia." *The Review of Economics and Statistics*, 104(1): 157–165.
- Lukat, Justina, Jürgen Margraf, Rainer Lutz, William M. van der Veld, and Eni S. Becker.** 2016. "Psychometric Properties of the Positive Mental Health Scale (PMH-scale)." *BMC Psychology*, 4(1): 8.
- Mani, A., S. Mullainathan, E. Shafir, and J. Zhao.** 2013. "Poverty Impedes Cognitive Function." *Science*, 341(6149): 976–980.
- Mankiw, N. Gregory.** 2000. "The Savers-Spenders Theory of Fiscal Policy." *American Economic Review*, 90(2): 120–125.
- Martinelli, César, and Susan Wendy Parker.** 2009. "Deception and Misreporting in a Social Program." *Journal of the European Economic Association*, 7(4): 886–908.
- McKenzie, David.** 2012. "Beyond Baseline and Follow-up: The Case for More T in Experiments." *Journal of Development Economics*, 99(2): 210–221.
- McNeill, Kristen, and Rachael Pierotti.** 2021. "Reason-Giving for Resistance: Obfuscation, Justification and Earmarking in Resisting Informal Financial Assistance." *Socio-Economic Review*.
- Miller, Candace M., Maxton Tsoka, and Kathryn Reichert.** 2011. "The Impact of the Social Cash Transfer Scheme on Food Security in Malawi." *Food Policy*, 36(2): 230–238.
- Milligan, Kevin, and Mark Stabile.** 2011. "Do Child Tax Benefits Affect the Well-Being of Children? Evidence from Canadian Child Benefit Expansions." *American Economic Journal: Economic Policy*, 3(3): 175–205.
- Misra, Kanishka, Vishal Singh, and Qianyun (Poppy) Zhang.** 2022. "Frontiers: Impact of Stay-at-Home-Orders and Cost-of-Living on Stimulus Response: Evidence from the CARES Act." *Marketing Science*, 41(2): 211–229.
- Moore, Jeffrey C., Linda L Stinson, and Edward J Welniak.** 2000. "Income Measurement Error in Surveys: A Review." *Journal of Official Statistics*, 16(4): 331–361.
- Mullainathan, Sendhil, and Eldar Shafir.** 2013. *Scarcity: Why Having Too Little Means So Much*. New York, NY: Macmillan.
- O'Brien, Rourke L.** 2012. "Depleting Capital? Race, Wealth and Informal Financial Assistance." *Social Forces*, 91(2): 375–396.
- O'Donnell, Michael, Amelia S. Dev, Stephen Antonoplis, et al.** 2021. "Empirical Audit and Review and an Assessment of Evidentiary Value in Research on the Psychological Consequences of Scarcity." *Proceedings of the National Academy of Sciences*, 118(44): e2103313118.
- Oswald, Andrew J., and Stephen Wu.** 2010. "Objective Confirmation of Subjective Measures of Human Well-Being: Evidence from the U.S.A." *Science*, 327(5965): 576–579.
- Perrig-Chiello, P., W. J. Perrig, and H. B. Stähelin.** 1999. "Health Control Beliefs in Old Age—Relationship with Subjective and Objective Health, and Health Behaviour." *Psychology, Health & Medicine*, 4(1): 83.
- Persaud, Navindra, Kevin E Thorpe, Michael Bedard, Stephen W Hwang, Andrew Pinto, Peter Jüni, and Bruno R da Costa.** 2021. "Cash Transfer during the COVID-19 Pandemic: A Multicentre, Randomised Controlled Trial." *Family Medicine and Community*

- Health*, 9(4): e001452.
- Pignatti, Clemente, and Zachary Parolin.** 2023. “The Effects of an Unconditional Cash Transfer on Mental Health in the United States.” *IZA Discussion Paper*, 16237.
- Pilkauskas, Natasha V., Brian A. Jacob, Elizabeth Rhodes, Katherine Richard, and H. Luke Shaefer.** 2023. “The COVID Cash Transfer Study: The Impacts of a One-Time Unconditional Cash Transfer on the Well-Being of Families Receiving SNAP in Twelve States.” *Journal of Policy Analysis and Management*, 42(3): 771–795.
- Portes, Alejandro.** 1998. “Social Capital: Its Origins and Applications in Modern Sociology.” *Annual Review of Sociology*, 25.
- Präg, Patrick, Nina-Sophie Fritsch, and Lindsay Richards.** 2022. “Intragenerational Social Mobility and Well-being in Great Britain: A Biomarker Approach.” *Social Forces*.
- Price, David J, and Jae Song.** 2018. “The Long-Term Effects of Cash Assistance.” *Princeton University Industrial Relations Section working paper*, 621.
- Richterman, Aaron, Christophe Millien, Elizabeth F. Bair, Gregory Jerome, Jean Christophe Dimitri Surrin, Jere R. Behrman, and Harsha Thirumurthy.** 2023. “The Effects of Cash Transfers on Adult and Child Mortality in Low- and Middle-Income Countries.” *Nature*, 1–8.
- Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel.** 2020. “Poverty, Depression, and Anxiety: Causal Evidence and Mechanisms.” *Science*, 370(1289).
- Robertson, Laura, Phyllis Mushati, Jeffrey W Eaton, et al.** 2013. “Effects of Unconditional and Conditional Cash Transfers on Child Health and Development in Zimbabwe: A Cluster-Randomised Trial.” *The Lancet*, 381(9874): 1283–1292.
- Royle, Jane, and Nadina B. Lincoln.** 2008. “The Everyday Memory Questionnaire – Revised: Development of a 13-Item Scale.” *Disability and Rehabilitation*, 30(2): 114–121.
- Ruggeri, Kai, Amma Panin, Milica Vdovic, et al.** 2022. “The Globalizability of Temporal Discounting.” *Nature Human Behaviour*, 1–12.
- Salkind, Neil J., and Ron Haskins.** 1982. “Negative Income Tax: The Impact on Children from Low-Income Families.” *Journal of Family Issues*, 3(2): 165–180.
- Schilbach, Frank, Heather Schofield, and Sendhil Mullainathan.** 2016. “The Psychological Lives of the Poor.” *American Economic Review*, 106(5): 435–440.
- Sedlmayr, Richard, Anuj Shah, and Munshi Sulaiman.** 2020. “Cash-plus: Poverty Impacts of Alternative Transfer-Based Approaches.” *Journal of Development Economics*, 144: 102418.
- Shah, Anuj K., Jiaying Zhao, Sendhil Mullainathan, and Eldar Shafir.** 2018. “Money in the Mental Lives of the Poor.” *Social Cognition*, 36(1): 4–19.
- Shah, Anuj K., Sendhil Mullainathan, and Eldar Shafir.** 2012. “Some Consequences of Having Too Little.” *Science*, 338(6107): 682–685.
- Shah, Anuj K., Sendhil Mullainathan, and Eldar Shafir.** 2019. “An Exercise in Self-Replication: Replicating Shah, Mullainathan, and Shafir (2012).” *Journal of Economic Psychology*, 75: 102127.
- Silver, David, and Jonathan Zhang.** 2023. “Invisible Wounds: Health and Well-Being Impacts of Mental Disorder Disability Compensation on Veterans.” *National Bureau of Economic Research*, w29877.
- Simes, R. John.** 1986. “An Improved Bonferroni Procedure for Multiple Tests of Significance.” *Biometrika*, 73(3): 751–754.
- Sorokin, Pitirim A.** 1959. *Social and Cultural Mobility*. Glencoe, IL:Free Press.
- Stango, Victor, and Jonathan Zinman.** 2009. “Exponential Growth Bias and Household Finance.” *The Journal of Finance*, 64(6): 2807–2849.
- Stango, Victor, and Jonathan Zinman.** 2014. “Limited and Varying Consumer Attention:

- Evidence from Shocks to the Salience of Bank Overdraft Fees.” *Review of Financial Studies*, 27(4): 990–1030.
- Sugden, Robert.** 1985. “Regret, Recrimination and Rationality.” *Theory and Decision*, 19: 77–99.
- Szaszi, Barnabas, Bence Palfi, Gabor Neszveda, Aikaterini Taka, Péter Szécsi, Christopher Blattman, Julian C. Jamison, and Margaret Sheridan.** 2023. “Does Alleviating Poverty Increase Cognitive Performance? Short- and Long-Term Evidence from a Randomized Controlled Trial.” *Cortex*, 169: 81–94.
- Thaler, Richard H.** 1999. “Mental Accounting Matters.” *Journal of Behavioral Decision Making*, 12(3): 183–206.
- Toh, Ying Lei, and Thao Tran.** 2020. “How the COVID-19 Pandemic May Reshape the Digital Payments Landscape.” Federal Reserve Bank of Kansas City.
- Troller-Renfree, Sonya V., Molly A. Costanzo, Greg J. Duncan, et al.** 2022. “The Impact of a Poverty Reduction Intervention on Infant Brain Activity.” *Proceedings of the National Academy of Sciences*, 119(5).
- United States Department of Agriculture.** 2012. “U.S. Household Food Security Survey Module.” 12.
- Watson, Brett, Mouhcine Guettabi, and Matthew N Reimer.** 2019. “Universal Cash Transfers Reduce Childhood Obesity Rates.” *SSRN Electronic Journal*.
- Westfall, Peter H., and James F. Troendle.** 2008. “Multiple Testing with Minimal Assumptions.” *Biometrical Journal. Biometrische Zeitschrift*, 50(5): 745–755.
- Westfall, Peter H., and S. Stanley Young.** 1993. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. Hoboken, NJ:John Wiley & Sons.
- Widerquist, Karl.** 2005. “A Failure to Communicate: What (If Anything) Can We Learn from the Negative Income Tax Experiments?” *The Journal of Socio-Economics*, 34(1): 49–81.
- Wollburg, Clara, Janina Isabel Steinert, Aaron Reeves, and Elizabeth Nye.** 2023. “Do Cash Transfers Alleviate Common Mental Disorders in Low- and Middle-Income Countries? A Systematic Review and Meta-Analysis.” *PLOS ONE*, 18(2): e0281283.
- WorldBank.** 2015. “World - Global Financial Inclusion (Global Findex) Database 2014.”
- Yoo, Paul Y., Greg J. Duncan, Katherine Magnuson, et al.** 2022. “Unconditional Cash Transfers and Maternal Substance Use: Findings from a Randomized Control Trial of Low-Income Mothers with Infants in the U.S.” *BMC Public Health*, 22(1): 897.

How Effective Is (More) Money? Randomizing Unconditional Cash Transfer Amounts in the US*

Online Appendix

Ania Jaroszewicz¹, Oliver P. Hauser^{2,3}, Jon M. Jachimowicz⁴, and Julian Jamison^{2,5}

¹University of California, San Diego

²University of Exeter

³Harvard University

⁴Harvard Business School

⁵University of Oxford

July 20, 2024

*Jaroszewicz (corresponding author): aniaj@ucsd.edu. Hauser: o.hauser@exeter.ac.uk. Jachimowicz: jjachimowicz@hbs.edu. Jamison: J.Jamison@exeter.ac.uk. This paper was previously circulated under the title “Cash Can Make Its Absence Felt: Randomizing Unconditional Cash Transfer Amounts in the US.” This RCT was registered as AEARCTR-0006149.

Table of Contents

Appendices	3
A Prediction study	4
A.1 Methods	4
A.2 Results	5
B Main study: Data details	8
B.1 Index construction	8
B.2 Data preparation for survey responses	10
B.3 Sending and receiving money data details	10
B.4 Bank account balance and transactions data details	11
B.5 Duplicates in bank accounts	11
B.6 Data preparation for bank account balances	12
B.7 Data preparation for financial transactions	13
B.8 Coding of intended money usage responses	14
C Main study: Additional attrition analyses	15
C.1 Correlates of responsiveness	15
C.2 Selective attrition test	16
C.3 When is missingness problematic?	16
C.4 Ways that missingness could have biased estimates downward	17
C.4.1 Story 1	18
C.4.2 Story 2	18
C.4.3 Story 3	19
C.4.4 Story 4	20
C.4.5 Story 5	20
C.4.6 Discussion	21
C.5 Extreme value bounds	21
C.6 Honaker-King multiple imputation	23
C.7 Discussion	24
D Main study: Additional possible mechanisms	25
D.1 Strategic responding	25
D.1.1 Experiment	25
D.1.2 Correlational survey responses	26
D.1.3 Comparing self-reports against bank data	28
D.1.4 Discussion	28
D.2 Reference dependence	28
D.2.1 Experiment	28
D.2.2 Correlational survey responses	29
D.2.3 Discussion	30
D.3 Spending behavior	30
D.4 Expectations of money amounts	31
D.5 Inference from money receipt	32
D.6 Declining social relationships	33
D.7 Attrition	33

E Main study: Other additional analyses	34
E.1 Participant expectations	34
E.2 Reported vs. administrative savings and income figures	34
E.2.1 Methods	34
E.2.2 Results	35
E.3 Measuring treatment details	36
E.4 Effect of UCTs on transaction volume, debit volume, and net expenditures	37
E.5 Benchmarking survey result effect sizes	37
E.6 Objective and subjective outcome analysis details	38
E.7 Wave 6 narrower within-subjects analysis	39
F Prediction study: Additional figures	41
G Main study: Additional figures	44
H Main study: Additional tables	53
I Model	64
I.1 Introduction	64
I.2 Optimal d^*	66
I.2.1 Isoelastic case	66
I.3 Strategy	67
I.4 Optimal \bar{d}	67
I.5 Parametrization and simulations	67

A Prediction study

To provide credible priors on people’s beliefs about the effectiveness of unconditional cash transfers, we conducted a prediction study to obtain estimates from experts and laypeople about how effective the two cash treatments in the field experiment would be for the specific population, context, and outcomes.

A.1 Methods

Data collection for the prediction study was concurrent with that of the main study, taking place from November 2020 to January 2021. When we launched the prediction study, only 1% of the post-treatment surveys from the main study had been completed. Participants were recruited through a range of channels: the Social Science Prediction Platform (DellaVigna, Pope and Vivaldi, 2019), the Harvard Kennedy School Behavioral Insights Group monthly newsletter, our personal networks, and a series of academic listservs (specifically, the Economic Science Association, the Society for Judgment and Decision-Making, the European Association for Decision-Making, the Society for Personality and Social Psychology, and the Academy of Management). These channels brought in the bulk of the “expert” participants.¹ On behalf of each participant recruited through these channels, we donated a fixed amount to one of three charities of the participant’s choosing. Our layperson participants were primarily recruited through the recruitment platform lu.cid. These participants were representative of the US population according to the following 2019 Census rates: gender, age, race, geographic region in the US (Midwest, Northeast, South, West), household income, education, and political preferences (Snowberg and Yariv, 2021). Participants recruited through lu.cid received a flat participation payment.

All participants received a brief description of the main study, including the study timeline, the treatment groups, our projected participant numbers (including expected attrition), and baseline characteristics for the first 1,476 participants (those for whom we had baseline data at the time of launching the prediction study). Participants recruited through lu.cid (but not the other channels) were asked comprehension check questions about this information; they were corrected if they answered one or more questions incorrectly. All prediction study participants were also given a link to the preregistration page for the main study, which included our preanalysis plan and more details. See the Social Science Registry (socialscienceregistry.org/trials/6149) for the survey materials.

After receiving information on the main study design, participants were informed that their goal would be to predict how effective each treatment (\$500 and \$2,000) would be at changing people’s outcomes at three time points: one week after cash receipt (t2), six weeks after cash receipt (t3), and 15 weeks after cash receipt (t4). They were given a brief tutorial on effect sizes, a link to a one-page layperson description with optional additional information on effect sizes (also posted on the Social Science Registry #6149), and two practice problems unrelated to our study. Predictions were incentivized: they were told that after the main study concluded, we would randomly pick

¹The idea of using an expert sample is inspired by DellaVigna and Pope (2018).

one of the predictions participants made and the three people whose answers were closest to the true value would get a \$100 gift card.

Participants then proceeded to predict the effects for each of the four indices, which were presented in random order. For each index, they were asked to predict the effect sizes for each of the two treatment groups at each of the three post-treatment time points. Participants were given a summary of the questions in each index and, again, a link to the preregistration page where the exact survey materials were provided. Predictions were restricted within a range of -1 to +1 *SDs*. (Five participants took an older version of the survey that did not have this restriction and entered at least one value outside these bounds. In the analyses, these predictions are censored at -1 or +1 as needed. However, all results are robust to excluding these individuals, as well.)

One potential concern with eliciting predictions about the indices is that they may be somewhat difficult to understand. Thus, we also included a second set of prediction prompts for an individual representative question from each index. For the financial index, this question was predicting the main study participants' agreement with the statement, "I can currently meet my and my household's basic needs"; for the psychological index, it was agreement with the statement, "Over the past week, I have been in a good emotional condition"; for the cognitive capacity index, this was participants' Raven's matrices score; and for the health index, this was whether participants felt that their overall physical health in the past week was poor, fair, good, very good, or excellent. All participants recruited through lu.cid completed these individual question predictions. Participants not recruited through lu.cid were asked whether they would like to do these predictions or skip to the end of the study; as a thank you for opting into doing the additional predictions, we increased their charity donation. Predictions for this second set of prompts were incentivized through a second accuracy contest to win an additional \$100 gift card.

We drop two responses with a duplicate email address and four responses where people indicated having taken the survey before. We are left with 1,448 participants in total. Of these, 914 were recruited through the lu.cid platform and the remainder were recruited through the other channels. The "layperson" and "expert" division we use in the analyses largely, though not perfectly, aligns with whether participants were recruited through lu.cid or the other channels, respectively. Among those recruited through the non-lu.cid channels, 79% opted into completing the additional predictions on the individual questions.

A.2 Results

In the analyses below, we divide the sample into "experts" (people who identified themselves as a researcher or policymaker; $N = 477$) and "laypeople" (people who did not identify themselves as such; $N = 971$). The left side of Appendix Figure F.1 plots the prediction study results for the indices. On average, participants predicted positive effects of both treatments on all four indices at all three post-treatment time points. Specifically, for each index at each time point, they predicted that the \$500 group would fare better than the Control group (one-sample two-tailed t-test for each index at each time period, measuring the difference from 0; all $p < 0.001$) and that the \$2,000

group would fare better than the \$500 group (separate OLS regression for each index for each time period with a dummy variable for treatment group using robust *SEs* clustered at the participant level; all $p < 0.001$). Using the standard Cohen categorization of effect sizes (Cohen, 1988), we can describe the predicted effect sizes as spanning from small to large. The largest average predicted effect size was for the \$2,000 group psychological index at t2 (0.65 *SDs*), while the smallest was for the \$500 group financial index at t4 (0.16 *SDs*). Only 2% of the expert predictions were negative.

In addition, participants predicted that, for each treatment group and each index, the effects would be most positive at t2, weaker at t3 (separate OLS regressions for each index for each treatment group with a dummy variable for time period using robust *SEs* clustered at the participant level; all $p < 0.001$), and weaker still at t4 (same specification; all $p < 0.001$). That is, both laypeople and experts believed that the effects would diminish over time.

Finally, for each treatment group, each index, and each time period, experts predicted smaller effect sizes than laypeople, although these differences were not always statistically significant (separate OLS regressions for each index for each treatment group for each time period with a dummy variable for expert using robust *SEs*; all $p \leq 0.144$). It is unclear whether these gaps are generated by a true difference in subjective evaluations of the effectiveness of cash or a difference in familiarity with effect size distributions.

Importantly, we repeat the analyses above for the cognitively simpler individual question predictions and find that the results are qualitatively similar, suggesting that the results are not driven by a misunderstanding of the indices. Specifically, all predictions were positive: participants anticipated that both cash amounts would improve responses to all four individual questions at all three time points. As with the indices, they predicted that the \$500 group would have better individual questions outcomes than Control (one-sample two-tailed t-test for each index at each time period, measuring difference from 0: all $p < 0.001$), and that the \$2,000 group would fare better than the \$500 group (separate OLS regression for each index for each time period with a dummy variable for treatment group using robust *SE* clustered at the participant level; all $p < 0.001$). The predicted effect sizes for the individual questions spanned a similar range as they did for the indices (0.17 to 0.70 *SDs*). The second main finding also holds: participants predicted that the effects would be strongest immediately after cash receipt and weaken over time. For each treatment group and each index, participants predicted that the effects would be more positive at t2 than t3 (separate OLS regressions for each index for each treatment group with a dummy variable for time period using robust *SEs* clustered at the participant level; all $p < 0.001$), and more positive at t3 than t4 (same analysis; all $p < 0.001$). Finally, just as with the indices, the experts also predicted smaller effect sizes for the individual questions, relative to laypeople (separate OLS regressions for each index for each treatment group for each time period with a dummy variable for expert using robust *SE*; all $p < 0.001$, except for the financial index \$500 group effect at t4: $p = 0.319$). These results suggest that the (mis)predictions are not driven by a misunderstanding of the indices. See the right side of Appendix Figure F.1.

In sum, both experts and laypeople expected positive effects of receiving (more) money. As

described at length in the main manuscript, these predictions do not correspond to our actual findings. Appendix Figure F.2 compares predictors' estimates of cash effectiveness with the true effectiveness, revealing an overestimation of 0.2 to 0.7 *SDs*, depending on the treatment group, index, and time period.

B Main study: Data details

B.1 Index construction

The four tables below describe the variables we used in the indices as part of our confirmatory hypotheses. See the study materials at socialscienceregistry.org/trials/6149 for the precise questions used. All outcomes below were measured in the baseline survey and at each time point thereafter, except for the Raven's matrices, which were measured only after the intervention.

Financial well-being

Variable name	Additional description
EMPLOYMENT	Binary indicator for whether the participant checked off any of the 3 “employed” options.
WORKPERFORMANCE	—
WORKSATISFACTION	—
SAVINGS	Winsorized at 90% (censoring values on the top and bottom 5%).
EARNEDINCOME	Winsorized at 90% (censoring values on the top and bottom 5%).
FINWELLBEING1, FINWELLBEING2, FINWELLBEING3, MEETNEEDS, MEETWANTS	Composite measure of all questions. FINWELLBEING2 and FINWELLBEING3 are reverse coded. Rescaled so all are on the same scale.
CREDITCONSTRAINTALLEV	—

Psychological well-being

Variable name	Additional description
AGENCY1, AGENCY2	Composite measure of all questions. AGENCY2 is reverse coded.
BESTLIFE, POSMENTALHEALTH1, POSMENTALHEALTH2, POSMENTALHEALTH3, HAPPY	Composite measure of all questions. Rescaled so all are on the same scale.
ANXIOUS, LONELY, DEPRESSION1, DEPRESSION2, DEPRESSION3, DEPRESSION4, DEPRESSION5, DEPRESSION6, DEPRESSION7, DEPRESSION8	Composite measure of all questions. All reverse coded. Rescaled so all are on the same scale.

Cognitive capacity

Variable name	Additional description
RAVENS	Score for the entire set of nine questions. Collected at every timepoint except t1.
MONEYMIND[t]A, MONEYMIND[t]B	The extent to which the participant would think about the cost-related statements, reverse coded and averaged across the two scenarios presented at each timepoint. The cost-related statements for each scenario are marked in the study materials.
MEMORY1, MEMORY2, MEMORY3, MEMORY4	Composite measure of all questions; reverse coded.

Physical health

Variable name	Additional description
HEALTH	—
SLEEP	—
FOODSECURITY	—
NUTRITION	—
EXERCISE	—

To construct the indices, we follow the approach detailed in Anderson (2008). All outcomes were first oriented such that higher values were “better” (i.e., indicated higher participant well-being). Next, we transformed the outcomes by demeaning them and dividing each one by the Control group standard deviation at each time point. This allowed us to compare the outcomes on a common scale. After standardizing the effect sizes, we created a weighted average of these transformed outcomes for each individual for each measure in each domain. The weight of each input was equal to the sum of the row entries in the inverted covariance matrix in each domain. This weighted average was used in our analyses.

As noted in Anderson (2008), with this procedure, the final outcome measure ignores missing values. Outcomes within a given domain that are highly correlated with one another receive relatively little weight within the index. In contrast, outcomes that are not highly correlated (and thus provide additional information) receive comparatively more weight.

The “money on the mind” measure used in the cognitive capacity index follows the directionality of the original Shah et al. (2018) work, with less thinking about money being “better” in the sense that it occupies less of a person’s mental bandwidth. However, as the model highlights, there

may be cases in which thinking about one’s finances is positive, if e.g. it allows people to address problems directly rather than avoiding them.

B.2 Data preparation for survey responses

In the survey response dataset, we drop: (i) t1 responses that were not completed on time and thus did not result in the participant becoming randomized (2,847 responses); (ii) post-treatment survey responses that were not begun within 30 days of when the survey was sent (847 responses); (iii) responses where there were errors in the process, specifically a survey link problem that prevented people from seeing the correct survey (28 responses), responses from people who were not randomized despite meeting the enrollment criteria (seven responses), and a person who was accidentally enrolled twice (four responses). In addition, in cases in which a participant began more than one survey within the 30 day window for a given t time period, we employ the following rules: keep the most complete survey response; if there are two or more surveys that are equally complete, keep the one that was started the earliest. Using these rules, we (iv) drop 2,810 responses and are left with at most one survey response per participant per time period. Our final dataset includes 16,747 responses, 15,483 of which are complete (meaning that participants reached the last page of the survey). Although participants were allowed to skip any questions they did not want to answer, and we retain responses that were begun but not completed, all 16,747 responses contain at least one of our preregistered dependent measures.

B.3 Sending and receiving money data details

The money-sending dataset from the non-profit organization includes payments they made to participants spanning from four months before our study began (April 1, 2020) to nine months afterwards (February 11, 2022). The overwhelming majority (97%) of payments made during the study window were related to our study. However, all payments regardless of source or reason are included in our analyses in Section 3.2 of the main text. We do not analyze data on payments made after the study concluded (i.e., after our dependent measures are collected). Whether the payment was made onto the online platform (requiring a participant to “pull” the money) or directly into the participant’s external account (involving a “push”) was determined by the non-profit organization’s internal workings, and in particular the funding account from which the money was coming and how the payment was cataloged internally.

The money-receipt dataset from the non-profit spans the same time period as the money-sent dataset and similarly tracks both money that is and is not related to our study. The amount of money received is in part a function of how the money was sent. If the non-profit pushed the payment directly to the participant’s external account, the amount received simply corresponds to the amount that was sent. If the non-profit instead placed the payment on their online platform for the participant to proactively pull, the amount received was determined by the participant. For instance, a participant receiving a \$500 UCT on the online platform could have decided to first

only pull \$400 and then pull the remaining \$100 at a later date. Alternatively, that participant could have waited to accrue two additional survey payments before pulling all \$540 at once.

B.4 Bank account balance and transactions data details

The amount of time for which we observe participants’ bank balance data varies as a function of when the participant provided access and what their financial institution’s policies were. After cleaning, we observe 3,349 bank accounts across 2,366 (45%) participants, totaling 357,134 person-bank account balance-days. The earliest observed date is February 5, 2019; the latest is July 19, 2021. For 95% of the participants who opted into providing their bank account information (i.e., for 43% of the total study sample), we observe bank data for at least a fraction of the time that they were enrolled in the study. For the remaining 5%, we observe their data outside the trial period but not during it. The average span of observed days is 158; the median is 171. We observe both current and available bank balances, but our analyses primarily focus on the former. Because bank account balances are typically observed as one “snapshot” per bank account per day, the data are uniformly distributed across days of the week and month.

The financial transactions dataset includes the same participants as the bank account balance dataset and is for the same temporal windows.

B.5 Duplicates in bank accounts

Both the bank account balance data and financial transaction data include duplicates. At least one source of these duplicates relates to the fact that participants were able to link the same account multiple times without the non-profit’s system stopping the participants or identifying that the accounts had already been linked. Unfortunately, each linkage resulted in a unique masked bank account ID number, even if the linkage was not to a unique bank account. In this section, we describe our process for uncovering these duplicates.

Because the two datasets share some information—namely, the masked bank account IDs—we can use the datasets together to gain some insights. We begin by first identifying suspicious bank accounts just using the bank account snapshots dataset, then identifying suspicious bank accounts just using the financial transactions dataset, and finally by identifying the overlap in these flags—that is, which bank accounts appeared to be suspicious using the bank account snapshot criteria only, which accounts appeared suspicious using the financial transaction criteria only, which accounts appeared suspicious using both criteria, and which did not appear suspicious using either criterion.

Beginning with the bank account snapshot dataset, we flag snapshots that are duplicates based on the participant’s unique ID number, the date on which the bank balance was observed, the current balance that a given bank account ID had for a given day,² and the available balance a

²In the rare case in which there were multiple observations for a given bank account ID on a single day, we first took the average of all observations for a given bank account ID on that day. In these cases, the variable used was the average balance for that day. See Appendix Section B.6.

given bank account ID had for a given day.³ Any observations that were duplicates on all these variables yet had different bank account IDs were flagged as potentially “suspicious snapshots” (31,877 out of 386,135, or 8.26% of snapshots). If a bank account ID ever has one or more “suspicious snapshots” (using the definition above), they are flagged as being potentially a “suspicious account by snapshot.” This criteria flags 439 out of 3,056 accounts (12.56%).

Next, we turn to the transactions dataset and repeat the exercise. Again we begin by identifying duplicates by the participant’s unique ID; the listed date of the transaction; the dollar amount of the transaction; the name of the store, service, or action (e.g., “Kroger,” “Cash Deposit,” “Visa Money Transfer”); and the transaction category name (e.g., “Shops,” “Transfer,” “Food and Drink”). We then identify which of these duplicates involve different bank account IDs. This ensures that we do not flag, for instance, transactions in which a person makes two \$100 ATM withdrawals from the same checking account in the same day, or buys three songs on iTunes using the same debit card on the same day. Using these criteria, we find that 100,520 out of 954,724 transactions (10.53%) are potentially “suspicious transactions.” In a parallel fashion to what we did with the snapshots, we then flag a bank account ID as being a potentially “suspicious account by transaction” if they ever had one or more transactions flagged as potentially suspicious.

Next, we can use the different “suspicious account” flags to identify where there is overlap in the two data sources: 79.5% of accounts are non-suspicious by both data sources, 10.5% are flagged as potentially suspicious by both data sources, and the remaining accounts are flagged as suspicious using one data source but not the other.

B.6 Data preparation for bank account balances

We begin with 391,408 bank account balance snapshots and clean the data in three ways. We first address the fact that occasionally we have more than one observation from a single bank account on a given day. Because our unit of analysis is on the day level, we first create an average bank balance for a given bank account ID in a given day. This averages in 5,273 snapshots (1.3%) of our observed snapshots, leaving us with 386,135 observations.

We next address the duplicates (see Appendix Section B.5). Our primary deduplication approach drops the individual “suspicious snapshots” from analysis (dropping 31,877 out of 386,135, or 8.26%, of observations). We believe that this approach most appropriately balances caution with providing a realistic and complete view of the bank data. However, we note that all of the results are robust to (i) not dropping any observations (i.e., taking the data at face value); (ii) dropping entire accounts that are flagged as suspicious by both the snapshots and transactions dataset metrics (this would drop 41,244 out of 386,135 accounts, or 10.68% of observations); and (iii) dropping entire accounts that are flagged as suspicious using either the transactions or snapshots dataset metrics (this would drop 57,318 out of 386,135 accounts, or 14.84% of observations).

Finally, if a person has multiple observed balances on a given day that come from seemingly truly different bank account IDs, then we sum those balances to create a “total current balance by

³See above footnote.

person by day” variable. This is because we are interested in cumulative wealth rather than the amounts in separate bank accounts per se. In our final dataset, we have only one cumulative bank balance per person per day: a total of 357,134 observations.

B.7 Data preparation for financial transactions

Our raw and uncleaned financial transactions dataset contains 954,724 transactions. To prepare the dataset for analysis, for our primary analyses we drop any payments from the non-profit that occurred on or after that participant’s trial start date (4,328 observations were dropped, leaving 950,396). This is to ensure that our analyses capture the effects of the money rather than the treatment itself. (These observations are kept for the analysis in which we compare self-reported income to administratively-observed bank inflows [Appendix Section E.2].)

We also deduplicate the transactions (see Section B.5). Our primary analysis drops potentially suspicious transactions, resulting in the removal of exactly 100,000 transactions (10.5% of observations) and leaving 850,396. However, like with the snapshots data, all of our conclusions are robust to using the alternative deduplication methods: (i) not dropping any observations; (ii) dropping entire accounts that are flagged as suspicious by both the snapshots and transactions dataset metrics (this would drop 111,827 transactions, or 11.8% of observations); and (iii) dropping entire accounts that are flagged as suspicious using either the transactions or snapshots dataset metrics (this would drop 152,438 transactions, or 16.0% of observations).

We note two features of this dataset. First, we observe different spans of data for different participants. For instance, for some we have access to their account for three months while for others we have access for six months. Second, a lack of a transaction on a given day means different things depending on what other transactions we observe. If, for instance, we observe a transaction on January 1 and January 3 but nothing on January 2, we can infer that the participant did not spend or receive money on January 2 using that account, and thus that the lack of a transaction can meaningfully be interpreted as a “0.” On the other hand, if we observe transactions on January 1 and 3 but nothing after that, then it is not clear whether the participant spent or received any money on January 4. Thus, the lack of that transaction should be interpreted as true missingness. Because of these two features, our primary operationalization of spending, net expenditures, the number of debits, and the number of transactions uses daily averages.

To construct daily averages, we first calculate the span of time for which we have access to a participant’s account in each of three time periods: before the UCT, after the UCT during the trial, and after the trial. If there is any observed transaction before the UCT date, the span before the trial is simply the first transaction date until the UCT date (if we observe any transactions on or after the UCT date) or until the last observed date (if the last observed date is before the UCT date). If there are no observed dates before the UCT date, then there is no span before the trial. The span after the UCT during the trial is the UCT date to the last day the t4 survey was open (if the first observed date is on or before the UCT date and the last observed date is on or after the last day the t4 survey was open). If the first observed transaction date was later, or the last

observed date was earlier, the span gets truncated accordingly. Finally, the span after the trial is the last day the t4 survey was open (if the first observed transaction date is on or before then) or the first observed date (if the first observed date is after the last day the t4 survey was open) until the last observed date (if the last observed date is after the last day the t4 survey was open). For narrower set-width windows (in particular, the two week periods following the UCT), we treat those spans as they are (14 days long). We then calculate daily spending, net expenditures (debits minus credits), the number of debits, and the number of transactions over each window of interest and divide by the inferred span of days to get daily averages.

B.8 Coding of intended money usage responses

Participants who withdrew money from the online platform were asked how they planned on using the money. Here, we describe the coding for these participant-generated responses (8,438 responses). Two research assistants (Rater 1 and Rater 2) were asked to categorize each response into one or more of 22 named categories (e.g., Housing, Transportation, Savings, General Debt; see Figure 4 in the main text for all 22 named categories) or an “Other” category. The dataset also includes 35 observations with withdrawal reasons generated by the non-profit organization, for a total of 8,473 observations. For completeness, these organization-generated withdrawals are also included in the analyses below, categorized as “Other.”

To calculate Cohen’s kappa (which requires mutually exclusive categories), we transform the 22 named categories and the “Other” category (23 categories in total) into new composite categories that are combinations of the original ones. For instance, if the rater indicated that a response mentioned both groceries and housing, we create a composite category called “Groceries+Housing.” Across all participant responses, we create 190 unique categories for Rater 1 and 177 unique categories for Rater 2. From these composite categories, we calculate a Cohen’s kappa of 0.8573 ($SE = 0.0036$; based on 87.23% agreement and 10.50% expected agreement), indicating very high agreement (Landis and Koch, 1977). After identifying responses where there was disagreement, the raters re-evaluated those responses and settled on a single coding, which constitutes our final dataset.

C Main study: Additional attrition analyses

C.1 Correlates of responsiveness

Appendix Table H.7 details the predictors of responsiveness. Models 1 to 3 use a binary indicator for whether a participant provided any post-treatment survey data, while Models 4 to 6 examine the number of observed post-treatment surveys.

Treatment group The likelihood of responding to any post-treatment survey was lower among the Control group (80%) than the \$500 group (90%; difference from Control: $p < 0.001$) and the \$2,000 group (88%; difference from Control: $p < 0.001$). There are no differences between the two cash groups ($p = 0.095$). We observe a similar pattern when using the continuous measure of responsiveness, with the Control group on average completing 1.8 post-treatment surveys, the \$500 group completing 2.3 (difference from Control: $p < 0.001$), and the \$2,000 group completing 2.2 (difference from Control: $p < 0.001$). Again, there are no differences between the two cash groups ($p = 0.608$).

Indices at baseline Importantly, none of the indices at baseline predict responsiveness using either the binary or continuous measures (all $p \geq 0.174$). See Appendix Table H.7.

Demographics and financial profile Because randomization and attrition occurred after baseline values were collected, we can use data from t1 to identify who left the study. We regress the binary and continuous responsiveness measures on gender, age, a binary indicator for not being exclusively White, having completed more than high school, household size, a binary indicator for being a parent, a binary indicator for having a partner or spouse, unearned income at t1, and a binary indicator for being under the federal poverty line in 2019. The results reveal that older participants, participants more likely to be exclusively White, and participants living in smaller households were more likely to respond, and the other variables do not carry predictive power. See Appendix Table H.7.

Treatment group interacted with indices at baseline When we interact a binary indicator for being in either cash group with each of the indices at baseline, we find mixed results. For the financial index interaction, the coefficient is negative and statistically significant; for the psychological index interaction, the coefficient is positive and statistically significant for the continuous measure of responsiveness and null for the binary measure; and for the cognitive capacity and health index interactions, the coefficients are not significant. See Appendix Table H.8.

Bank data To further investigate the relationship between a person's financial profile and responsiveness, we regress our measures of responsiveness on six predictors from the bank account data: the number of bank accounts linked, average bank account balance before the UCT, daily amount spent before the UCT (i.e., debits), daily net expenditures before the UCT (i.e., debits

minus credits), daily number of debits before the UCT, and daily number of transactions (debits and credits) before the UCT. We find that those who opted to link more bank accounts were also substantially more likely to respond to the post-treatment surveys (binary measure of responsiveness: $\beta = 0.723, p < 0.001$; continuous measure of responsiveness: $\beta = 0.168, p < 0.001$). In addition, daily spending before the UCT predicted the continuous measure of responsiveness ($\beta = -0.002, p = 0.021$), but not the binary one ($\beta = -0.003, p = 0.291$). None of the other variables predicted responsiveness (all $p \geq 0.291$).

C.2 Selective attrition test

We conduct a “selective attrition rate test” to determine whether, conditional on response status, observable baseline outcomes differ across treatment groups. We are primarily concerned with internal validity for the respondent subpopulation. Following the guidance and code of Ghanem, Hirshleifer and Ortiz-Beccera (2023), we conduct a joint test of the equality of the baseline outcome distribution between respondents in the three treatment arms, as well as attritors in the three treatment arms. Our first test checks for differences in baseline values of the four indices.⁴ Reassuringly, we do not reject the null hypothesis of equality of baseline outcome distributions ($p = 0.129$). For completeness, we also repeat the exercise when including a rich set of covariates collected during the t1 survey: being female, age, being not exclusively White, having completed more than high school, household size, being a parent, having a partner or spouse, debt, unearned income, and being under the federal poverty line in 2019.⁵ Again, we find that we cannot reject the null ($p = 0.242$). These results suggest that even when using our rich baseline data, there is no evidence that the Control, \$500, and \$2,000 group respondents and attritors systematically differed in observable ways.

C.3 When is missingness problematic?

Recall that we found no evidence of (more) cash improving outcomes, neither between the Control and cash groups, nor between the two cash groups. The fact that the two cash groups had similar response rates suggests that the latter result (the lack of a positive treatment effect when moving from \$500 to \$2,000) is unlikely to be driven by attrition. However, it is possible that the differences in response rates between the Control and cash groups could have contributed to the former result.

Consider four theoretically distinct types of participants: “always-reporters” (who always respond to our post-treatment surveys, regardless of their treatment group), “never-reporters” (who never respond to our post-treatment surveys, regardless of their treatment group), “if-treated reporters” (who respond to our post-treatment surveys if they are in one of the cash groups, but not otherwise), and “if-control reporters” (who respond to our post-treatment surveys if they are

⁴All participants have a baseline value for at least one index. However, nine participants have fewer than four baseline indices. For this analysis, these participants must be dropped.

⁵For this analysis, we must additionally drop any participants who are missing data for one or more of these variables (343 participants). We are thus left with $N = 4,891$.

in the control group, but not otherwise) (Gerber, 2012; Ghanem, Hirshleifer and Ortiz-Beccera, 2023). Naturally, it is impossible to tell which of the participants we observe in the cash groups are always-reporters versus if-treated reporters; and it is impossible to tell which of the participants we fail to observe from the Control group are never-reporters versus if-treated reporters. Nevertheless, it is useful to draw the conceptual distinctions. In the context of our attrition analyses, the if-treated reporters are the participants of greatest interest. For brevity, call the if-treated reporters in the cash groups (who are observed because they were treated) the “bonus” treatment group participants and the if-treated reporters in the Control group (who are not observed because they were not treated) the “conditionally-missing” Control group participants.

There are two possible states of the world. First, it is possible that the bonus treatment group participants on average had better outcomes (higher survey index values) than the Control group participants we observe. In this case, the differential attrition would bring up the average of the cash groups, making them look “artificially” better than Control. This would imply that the true (unobserved) effect of cash was lower or more negative than the data indicates it was.

Alternatively, it is possible that the bonus treatment group participants on average had worse outcomes (lower index values) than the Control group participants we observe. In this case, the differential attrition would make the cash groups on average look artificially worse than Control. The true (unobserved) effects of cash might be positive, but because of differential attrition, we would be concluding that they are not. Given the results of the experiment, this is the situation of concern.

A different way of thinking about this is to consider who the conditionally-missing Control group participants are. Again, there are two possible states of the world. If they tended to have better outcomes than the observed Control group participants, then their missingness would be suppressing the Control group average. Adding them back in would bring up the average of the Control group (to make it look even better than the cash groups), and therefore make our conclusions even more surprising. In the second state of the world, the conditionally-missing Control group participants we do not observe truly had worse outcomes than the Control group participants we observe. In this case, their missingness would bring up the average of the Control group, and potentially produce the pattern we observe: the Control group looks weakly better than the cash groups. This is the primary concern: that we are falsely casting the effects of cash as being non-positive when in fact they are positive.

C.4 Ways that missingness could have biased estimates downward

In this section, we lay out a few different constellations of assumptions and phenomena (“stories”) that would have to be true in order for the bonus treatment group participants to have worse outcomes than the Control group participants we observe and/or for the conditionally-missing Control group participants to have worse outcomes than the Control group participants we observe. That is, we identify ways that missingness could have biased our survey estimates downward. We then test the plausibility of each of those stories.

C.4.1 Story 1

One possibility is that, for whatever reason, the cash group participants who were initially comparatively worse off stayed in the trial, and/or the Control group participants who were initially comparatively well off stayed in the trial. To test this possibility, we regress the number of post-treatment surveys taken on the interaction between (eventual) treatment group randomization and index values at baseline (see Appendix Table H.8). Although cash group participants who were initially financially better off were less likely to respond to subsequent surveys, we observe the opposite pattern for the psychological index and no statistically significant relationship for the cognitive capacity or health index. Repeating the exercise using a binary indicator of having answered any post-treatment survey, we find similar results, except that the psychological index interaction loses significance. We interpret these results as suggesting that the evidence for this story is inconclusive.

C.4.2 Story 2

Recall that the study was run largely over the first year of the COVID-19 pandemic, a time of great financial and public health turbulence. If no participants had attrited, then successful randomization would have ensured that participants in the three treatment groups would have been roughly equally likely to live in areas with better versus worse macroeconomic and public health outcomes (treatment effects notwithstanding). However, attrition could have shifted the participant pool makeup such that the observed cash group participants lived in areas that were worse on these dimensions. This could happen if, for instance, reciprocity or trust encouraged cash group participants to respond to the surveys despite difficult circumstances in their area, thereby making the observed cash group participants disproportionately likely to live in areas with worse outcomes.

In this section, we test this possibility. Because we recruited participants from across the country and at different time points over seven months in the first year of the pandemic, there is substantial naturally occurring variation in macroeconomic and public health outcomes for the participants. We pull data from the Bureau of Labor Statistics on seasonally-adjusted unemployment rates, average number of hours worked per week, and average hourly earnings by month and state,⁶ as well as data from the Center for Disease Control and Prevention on COVID case and death rates by day and state (both are 7-day averages per 100,000 inhabitants).⁷ We then match these data to each participant for each post-treatment survey they were sent. For instance, if a participant living in Illinois was sent a t3 survey on January 24, 2021, we identify the seasonally-adjusted unemployment rate, average hours worked, and average hourly wages in Illinois in January 2021 and the COVID case and death rate in Illinois on that day.

Next, restricting our analyses to only those participants who answered a given time period survey, we test whether there is an association between being in a cash group and these five indicators for the time period, controlling for the indicator on the date the participant received their

⁶Data taken from <https://www.bls.gov/eag/home.htm> and <https://data.bls.gov/cgi-bin/dsrv?sm>

⁷Data taken from <https://covid.cdc.gov/covid-data-tracker>

t_1 survey. If the observed cash group participants tended to live in areas with worse employment and/or COVID outcomes than observed Control group participants, this could help explain why they looked unexpectedly bad on the survey outcomes. We find that, out of the 15 regressions we run (5 indicators \times 3 post-treatment time periods), only one shows a statistically significant relationship in the direction that would be consistent with this story (COVID death rate at t_2 , $\beta = 0.202$, $p = 0.014$). We conclude that it is unlikely that differential macroeconomic and public health outcomes contributed to the observed survey outcomes.

C.4.3 Story 3

In the third “story” we consider, four things must be true. (1) First, negative financial or health shocks occurring after we collected our baseline data must decrease responsiveness. This could happen if, for instance, participants are struggling so much with their lives that they do not want to or cannot be bothered with completing surveys. (2) Second, receiving a UCT must increase responsiveness to surveys, for instance, due to reciprocity or trust towards the experimenters, an increased subjective assessment that responding could result in more money, or simply by providing a sufficiently positive financial shock that can work in the opposite direction of (1). (3) Some fraction of cash participants experienced a negative financial shock after baseline but were also more likely to respond to the survey due to (2), and (2) was sufficient to convince a cash participant who otherwise would have attrited [in the absence of the effect of (2)] to instead respond. (4) Finally, for the group described in (3), the effect of the negative financial or health shock on pushing survey index outcomes “downward” (i.e., in making people look “worse” on indices) must be greater than the effect of reciprocity, trust, strategic responsiveness, or positive financial shock in pushing survey responses “upward.”

To identify the plausibility of this story, we test point (1). Although by definition we do not have survey data on post-treatment financial or other negative shocks to attrited participants, we can attempt to proxy for them using the local macroeconomic and public health data described in Appendix Section C.4.2. Across all participants, we separately regress responsiveness for each survey (logistic regression) on each of the macroeconomic and public health variables while controlling for those variables as they were on the date that the participant received their t_1 survey. We find very little evidence that worse macroeconomic and public health indicators are associated with lower likelihood of responding. Out of the 15 variables we test (5 indicators \times 3 time points), only one is statistically significantly associated with likelihood of responding in the direction that would be consistent with this story (average weekly hours worked at t_2 on responsiveness at t_2 ; $\beta = 0.275$, $p = 0.023$). Six are statistically significant in the opposite direction, and the remainder are null. We thus fail to find evidence consistent with the first requirement of this story, and so conclude that the story is unlikely to be at play in our study.

C.4.4 Story 4

The fourth story is a bit simpler than the last. We consider the possibility that simply due to busyness, disinterest, etc., there is a constant fraction of participants from each treatment arm who would have dropped out in the absence of treatment, but that the cash treatment compelled the if-treated cash group participants to instead respond. As mentioned in Section C.4.3, this could occur because of an increase in reciprocity or trust. To explain our primary results—that the cash group appeared to be doing similarly or worse than Control—this set of “bonus” cash group participants would need to have worse outcomes than Control on average.

Of course, we cannot identify who the “bonus” cash group participants are, and we lack data to test reciprocity towards the experimenter. However, we can test whether trust is positively or negatively associated with index values. At the end of the t4 survey, we asked participants to what extent they trusted that the experimenter would take the promised actions (e.g., provide timely payments, keep answers confidential). We find that the cash groups had somewhat higher trust estimates, though those differences did not reach statistical significance ($\beta_{\$500} = 0.095, p = 0.067$; $\beta_{\$2000} = 0.105, p = 0.110$). Importantly, however, trust had a null or positive, not negative, relationship with the index values (financial: $\beta = 0.040, p = 0.007$; psychological: $\beta = 0.050, p = 0.001$; cognitive capacity: $\beta = 0.011, p = 0.456$; health: $\beta = 0.013, p = 0.411$), even when including treatment group indicators (financial: $\beta = 0.042, p = 0.005$; psychological: $\beta = 0.052, p = 0.001$; cognitive capacity: $\beta = 0.013, p = 0.411$; health: $\beta = 0.056, p < 0.001$). These results are thus inconsistent with Story 4.

C.4.5 Story 5

Story 5 is just like Story 4 except, instead of trust/reciprocity, we test the possibility that the “bonus” cash group participants were more likely to believe that responding would result in getting more money. The data provide us with proxies for both pieces of data we need to calculate this expected value: subjective assessments of the probability and estimates of the amount of money that would be received, conditional on getting anything at all.

We begin with subjective assessments of the probability. In t4, participants were asked why some participants received one benefit and others got something else. These responses were coded, and participants who indicated that they thought the way they responded to the survey (either their answers conditional on responding or, more importantly for the analysis here, whether they responded at all) received a flag. We use this flag as a proxy for beliefs about whether responding could increase the likelihood of receiving subsequent funds. We find there are no differences across treatment groups in the prevalence of this flag. See Appendix Table H.12. Moreover, when we regress the four indices on this flag (for all post-treatment time periods with robust standard errors clustered at the participant level), we find no statistical relationship, regardless of whether the treatment group indicators are also included or not (all $p \geq 0.338$).

The second variable we use to proxy for these beliefs relates to the amount of money that could be received, conditional on it being received at all. In the t4 survey, we asked, “Suppose that

[non-profit organization] did give you more money. What is your BEST GUESS as to how much more you would receive?” This measure captures beliefs at the end of the t4 survey, and thus may not reflect beliefs several months earlier, but it is the best proxy we have. In this context, it is not surprising that participants in the cash groups believed that, if they were to receive any additional funds, the amount would be higher (relative to what the Control group believed) ($\beta_{\$500} = \163 , $\beta_{\$2000} = \470 , both $p < 0.001$). Moreover, these beliefs negatively predict the index post-treatment values (dividing the regressor by 100 for ease of interpretation, financial: $\beta = -0.012$, $p < 0.001$; psychological: $\beta = -0.011$, $p = 0.002$; cognitive capacity: $\beta = -0.010$, $p = 0.004$; health: $\beta = -0.012$, $p = 0.001$). The results are similar when including the treatment group indicators as additional regressors. Thus, while the effect sizes are small (each \$100 of additional money the participant expected is associated with a decrease of 1/100 of 1 point in index values), we do have evidence that these beliefs were more prevalent among the cash groups and were negatively associated with survey responses. Overall, the evidence for this pathway is mixed: no evidence for the first piece of the equation and supportive evidence for the second.

C.4.6 Discussion

In this section, we identified the situation in which attrition would be most problematic for our conclusions, then tested several possible stories that could produce that situation. Overall, we find very limited evidence for the first four stories, and some—but not fully conclusive—evidence for the final one.

C.5 Extreme value bounds

For the 38% of participants who responded to one or two post-treatment surveys, we employed the following imputation process. For each index, if the participant was missing a t2 value, we used their t3 value if it was available or their t4 value if t3 was not available. If they were missing a t4 value, we used their t3 value if it was available or their t2 value if t3 was not available. If they were missing a t3 value, we took the average of their t2 and t4 values, or just one of those two if they were not both available. After this imputation, we have complete post-treatment survey data for all but the 17% of participants who did not respond to any post-treatment surveys.

Following our preanalysis plan, we then calculated Lee bounds (Lee, 2009). We find that the lower bound coefficients range from -0.373 to -0.277 , depending on the index and treatment group, and upper bounds range from 0.046 to 0.253 ; see Appendix Table H.9. Results are similar if we instead treat the two cash groups jointly as the “treated” group and compare them to Control.

For completeness, we also calculate Horowitz-Manski bounds (Horowitz and Manski, 2000), which do not require monotonicity. To find the upper bounds, we replace each missing index value in the treatment group with the highest observed index value for that time period (for any of our treatment arms) and replace all missing index values in the Control group with the lowest observed index value. The lower bound is calculated by reversing the approach, such that the missing treatment group cells are replaced with the lowest observed index value and the missing Control

group cells are replaced with the highest observed value. We then re-run our primary analyses. As is often the case with extreme value bounds on variables that have a wide range of possible values (Gerber, 2012), we find fairly extreme bounds on both the lower (using our primary prespecified analysis: financial: $\beta_{\$500} = -1.27, \beta_{\$2000} = -1.30$; psychological: $\beta_{\$500} = -1.03, \beta_{\$2000} = -1.13$; cognitive capacity: $\beta_{\$500} = -1.25, \beta_{\$2000} = -1.34$; health: $\beta_{\$500} = -1.24, \beta_{\$2000} = -1.34$; difference from Control: all $p < 0.001$) and upper ends (financial: $\beta_{\$500} = 0.87, \beta_{\$2000} = 1.02$; psychological: $\beta_{\$500} = 0.93, \beta_{\$2000} = 0.95$; cognitive capacity: $\beta_{\$500} = 1.15, \beta_{\$2000} = 1.19$; health: $\beta_{\$500} = 0.91, \beta_{\$2000} = 0.94$; difference from Control: all $p < 0.001$). These figures suggest that it is at least theoretically possible that cash had positive effects on the indices.

To identify precisely when our conclusions would change, we invert and modify the exercise above, calculating how extremely positive the values would need to be for the missing cash group participants and how extremely negative the values would need to be for the missing Control group participants, in order to reverse our effect such that the cash groups on average would have more positive values than Control (Baird, McIntosh and Özler, 2019). Specifically, instead of using the most extreme values in our imputations, we impute the treatment arm's mean of an index at a given time point plus or minus a fraction of the index's arm-specific standard deviation at that time point. We vary the fraction amount by 0.01 *SDs* iteratively until the upper bound β coefficients no longer show a positive effect of cash for either the \$500 group or \$2,000 group. That is, we look for the largest *SD* fraction for each index for which neither treatment group upper bound coefficient is positive. These calculations allow us to see how extreme the missing participants' values could be such that we would still reach the same conclusion: that cash had no positive effect on the survey outcomes.

For the financial index, we find that the results still hold so long as the cash group participants' values were on average the mean financial index value plus 0.17 *SDs* or less, and the Control group participants' values were the mean minus 0.17 *SDs* or less—that is, so long as the gap between them around the mean is no larger than 0.34 *SDs* in the opposite direction of what we observe on average. For the psychological index, we find a maximum fraction of 0.22 *SDs* (total gap of 0.44 *SDs*); for the cognitive capacity index, 0.09 *SDs* (total gap of 0.18 *SDs*); and for the health index, 0.23 *SDs* (total gap of 0.46 *SDs*).

An alternative analysis could instead identify each index's largest *SD* fraction for which neither upper bound treatment group β coefficient is statistically significantly positive at an $\alpha = 0.05$ level (rather than identifying the largest fraction for which neither coefficient is positive at all). With this analysis, we find maximum values of 0.33 *SDs* for the financial index (total gap of 0.66 *SDs* in the opposite direction of what we observe on average), 0.33 *SDs* for the psychological index (total gap of 0.66 *SDs*), 0.21 *SDs* for the cognitive capacity index (total gap of 0.42 *SDs*) and 0.34 *SDs* for the health index (total gap of 0.68 *SDs*).

To put these numbers into context, we estimate the (non-causal) “effect” of living above versus below the median household income in the previous year on participants' baseline index values. Moving from below the median to above it is associated with an increase of 0.69 *SDs* for the

baseline financial index, 0.15 *SDs* for the psychological index, 0.09 *SDs* for the cognitive capacity index, and 0.18 *SDs* for the health index. Comparing these numbers to the gaps calculated above between the missing cash and missing Control group participants implies that, depending on the index, the gaps would need to be 0.96 to $4.67 \times$ the effect of moving from below the median income to above it in the previous year (financial index: $\frac{0.66}{0.69} = 0.96$, psychological index: $\frac{0.66}{0.15} = 4.40$, cognitive capacity: $\frac{0.42}{0.09} = 4.67$, health: $\frac{0.68}{0.18} = 3.78$).

We interpret these results as suggesting that it is possible that the true effect of cash was null on the survey outcomes and attrition made them appear negative in certain specifications, but it is highly unlikely that the true effect of cash was meaningfully positive for all the indices and attrition produced the observed effects.

C.6 Honaker-King multiple imputation

Following the methodology presented in Honaker and King (2010) (and per our preanalysis plan), we employ time-series cross-sectional imputations for the missing values. This method aims to incorporate all available information in the observed dataset to multiply impute missing values. In contrast to single imputation methods, which tend to overstate the certainty of coefficients and standard errors, multiple imputations provide more accurate estimates of the inferences' uncertainty because they take into account the variations across the multiple imputations for each missing value.

Honaker and King (2010) offers a detailed description of the approach; here, we describe it briefly. The algorithm makes two assumptions. First, it assumes that the dataset of dependent and explanatory variables used in the imputations and analyses are multivariate normal. To this end, we take the log of skewed variables such as debt and unearned income to normalize them. Second, the algorithm makes the usual assumption that the data are “missing at random” (but not “missing completely at random”—a more restrictive assumption). As described in Honaker and King (2010), the more information is included in the imputation model, the less stringent this assumption becomes. Thus, to help the algorithm, we include all variables from the survey, bank account balance, and financial transaction datasets that could potentially help with imputation, only removing open text responses, dates, redundant variables, nominal variables with more than 10 categories (which the algorithm has difficulty with; e.g., state of residence), and variables that are randomly assigned conditional on being observed (namely, variables pertaining to experiments embedded in the surveys).

To conduct the imputations, we used the Amelia package developed by the algorithm authors for the statistical software R. The algorithm combines the classic EM algorithm with bootstrapping to generate multiple complete datasets, each with different imputed values. We do not use any priors and follow the authors’ recommendation to create five datasets. Once the algorithm conducts the imputations, we combine the datasets and conduct the same analyses as we present in Section 4.1 of the main text, regressing each of the four indices on binary indicators for our treatment groups and, in some models, a range of covariates (see the “mi estimate” Stata commands and Rubin (2004)).

The results are reported in Appendix Table H.10. We find that the coefficients are somewhat less negative with this approach than they were without the imputations, but our conclusions remain qualitatively unchanged. We continue to primarily observe negative and statistically significant treatment effects for both cash groups on financial, psychological, and health survey outcomes (in two models, the \$2,000 group has negative, but not statistically significant, coefficients for the financial index) and primarily no effect of being in either cash group on the cognitive capacity index. In a robustness check, we redo the imputations with a second model that uses the four indices as lag and lead variables and find that the results are again qualitatively similar.

In summary, when the missing cells are filled using an approach specifically designed for this data structure, our conclusions are largely unchanged from when we did not impute values at all. These results are consistent with the interpretation that missingness did not play a major role in producing the observed non-positive effects of cash.

C.7 Discussion

Although we cannot fully rule out the possibility that differential attrition changed a meaningfully positive effect of cash on the survey indices into one that is indistinguishable from zero or negative, our analyses suggest that it is highly unlikely. First, the two cash groups had very similar response rates, suggesting that the lack of differences between them is unlikely (though not impossible) to have been caused by attrition.

Second, no index value at baseline predicts responsiveness, regardless of whether responsiveness is measured as a binary or continuous outcome. We also find inconsistent evidence regarding whether treatment group and index values at baseline interact to predict responsiveness.

Third, we identify a number of possible constellations of assumptions and phenomena that could have produced a pattern of missingness that in turn would have biased our treatment effectiveness estimates downward. Using a range of RCT and archival data, we test these possible “stories”—but find limited evidence for most of them in biasing our treatment estimates.

Fourth, we find that the gaps between the missing Control participants’ and missing cash group participants’ outcomes would have to be between 0.2 and 0.5 *SDs*, depending on the index—in the opposite direction of what is observed—for the effects of cash to be positive and not statistically significant. To observe statistically significant positive effects of cash, the gap would need to be between 0.4 and 0.7 *SDs*. The size of these effects would be possible, but fairly unlikely, particularly for the financial, psychological, and health indices. Our benchmarked estimates suggest that this would be equivalent to up to 4.7 times the “effect” of moving from below the median household income in the previous year to above the median.

Finally, when we multiply impute missing values using an approach specifically designed for time-series cross-sectional data, we find that our conclusions are largely unchanged: we continue to observe primarily negative effects of cash for the financial, psychological, and health survey outcomes, and primarily null effects of cash for the cognitive capacity outcome.

D Main study: Additional possible mechanisms

In this section, we describe the evidence for possible mechanisms that could explain the non-positive effects of (more) cash (in addition to the one we focus on in the main text). Our primary analyses use the full sample, but all results are robust to restricting the sample to only those participants who answered all the surveys.

D.1 Strategic responding

Although cash recipients were told that they had been randomly chosen to receive the UCTs, it is possible that participants nevertheless at some level believed that the non-profit organization gave money to people who they believed were in greatest need. If the cash groups also believed that they had more to gain financially from making themselves sound needier (relative to the Control group), they (but not the Control group) may have strategically distorted their responses “downward” (Moore, Stinson and Welniak, 2000; Martinelli and Parker, 2009; Baird and Özler, 2012; Beegle et al., 2012). We test this possibility in three ways: through an experiment added partway through the trial, through correlational survey responses, and by comparing self-reports of financial status with administrative bank data.

D.1.1 Experiment

In the experiment, we randomized 2,423 participants from Waves 5, 6, and 7 to one of two conditions. At the beginning of the t2 survey, the Informed group ($N = 1,236$) was told about the possibility of a new program with the partner non-profit organization. To heighten the saliency of the information, it was couched in a question about their potential interest in the program: “[Non-profit] may be enrolling people into new programs in the future. Would you potentially be interested in being part of a new [non-profit] program?” The Uninformed group ($N = 1,187$), on the other hand, did not receive this information. We reasoned that—if in fact participants were strategically attempting to make themselves sound needier—people informed of the potential new program would see larger benefits to distorting their responses downward, and so would appear to have worse outcomes than people who were not informed (assuming that the Uninformed participants were not already at a ceiling in terms of believing that they would get additional windfalls from the organization). We further reasoned that the effect might be larger with the cash group participants, who had just received a large payment and thus may have believed they had more to gain.

We find little evidence that being informed of the potential new program had an effect on survey responses. We first regress each of the four indices for the t2 survey on a binary indicator for being in the Informed group and find no main effect of the information treatment, relative to the Uninformed group (financial: $\beta = 0.070, p = 0.082$; psychological: $\beta = 0.049, p = 0.223$; cognitive capacity: $\beta = 0.014, p = 0.731$; health: $\beta = 0.050, p = 0.218$). We then regress each of the four indices (again for the t2 survey only) on an interaction of being in the Informed group and being

in one of the cash groups. For three of the indices, we find no statistically significant effect (all $p \geq 0.496$), and for the psychological index, we find a significant effect in the direction that would be consistent with the aforementioned theory ($\beta = -0.171, p = 0.034$).

Given that some Wave 6 participants took the t2 survey twice (see footnote 16 in the main text) and were independently randomized each time into either the Informed or Uninformed condition, we also construct a variable called EverInformed if the participant had ever been in the Informed condition. Rerunning the analyses above and only examining the results of Wave 6 participants in the second t2 survey, we now find no main effects (all $p \geq 0.125$) or interactions (all $p \geq 0.137$). The results are qualitatively similar when excluding Wave 6 participants entirely and only examining the other waves in t2.

Finally, it would be logical if the intervention had more of an effect like the one outlined above if the participant believed that their answers could affect how much money they received. At the end of the t4 survey, we asked participants, “Throughout this study, we asked you various questions about how you were doing financially (e.g., whether you could currently meet your and your household’s basic needs). How likely did you think it was that your answers to these questions would affect whether [non-profit] gave you money in the future?” Participants responded on a scale from (1) “Not at all likely” to (5) “Extremely likely.” We regress the four indices on an interaction of this variable and the indicator for being in the Informed condition and find no statistically significant interactions (all $p \geq 0.508$). We also regress the four indices on a triple interaction of being told about the potential new program, believing that one’s answers could affect how much money one received, and being in a cash group. Here we find no or positive associations between the interaction terms and the four indices—the direction that would be opposite of the story outlined above. All the results above are similar if we instead examine the effects on just the three financial well-being questions, which are the first set of more subjective items following the intervention. We thus find little evidence from the experiment to support the strategic distortion mechanism.

D.1.2 Correlational survey responses

In addition to the experiment, we also analyze correlational survey responses to test this mechanism. Beginning with the question on the perceived likelihood that one’s answers could affect how much money one received, we find that the Control, \$500 group, and \$2,000 group provided average responses of 2.49, 2.39 (OLS regression testing difference from Control: $p = 0.064$), and 2.53 (difference from Control: $p = 0.613$), respectively—indicating that overall participants thought it was somewhere between “a little likely” and “somewhat likely” that their responses would matter in this way. Responses to this question did not predict the financial ($\beta = 0.016, p = 0.212$) or psychological ($\beta = 0.013, p = 0.352$) index scores; negatively predicted the cognitive capacity index score ($\beta = -0.039, p = 0.002$), which would be consistent with the proposed theory; and positively predicted the health index score ($\beta = 0.030, p = 0.029$), which would be inconsistent with the proposed theory.

In t4, we also asked participants to describe why they thought that some people got certain

benefits and interventions in the study while others got something else. Note that this question asked participants about past interventions rather than future ones, and thus may not reflect their beliefs about what may affect future interventions. In addition, the question was asked at t4, and thus may not perfectly capture participants' beliefs at earlier time periods. Nevertheless, we believe it can provide some valuable insights.

Two research assistants (Rater 1 and Rater 2) blind to hypotheses and treatment groups coded the open-ended responses ($N = 2,374$) into one or more of eight pre-determined categories, including one for participant's survey answers and one for participant's characteristics or circumstances (e.g., financial need). Like with the responses on intended money usage (see Appendix Section B.8), we created combination categories for responses that were coded as falling into more than one category (e.g., a response that was coded as mentioning both the participant's survey answers and the researchers' interests was put into a new category called ParticipantSurveyAnswers&ResearcherInterest). This resulted in 38 distinct categories for both Rater 1 and Rater 2. Cohen's kappa on these 38 categories was 0.7182 ($SE=0.0096$; actual agreement=79.66%, expected agreement=27.84%).

After the raters settled their discrepancies and agreed on a single coding, we find that a substantial fraction (about 56%) of respondents reported that participants' characteristics or circumstances (including but not limited to financial circumstances) could have affected what they received in the study, though only 7% explicitly mentioned survey answers as a determinant. Moreover, there were few differences between treatment groups. The only statistically significant difference is that the cash groups were somewhat more likely to believe that the study allocations were determined by randomness (which is what cash participants were told in their UCT notification emails). See Appendix Table H.12 for a complete table of responses.

Like with the close-ended measure, we can also use these responses to assess how a reported belief that one's answers or circumstances can affect the amount of money received correlates with index outcomes. We create a new dummy variable that equals 1 if the participant stated that participants' survey answers and/or characteristics/circumstances (e.g., race, income, geographic region, or other information that could be communicated through survey responses) affected what they received in the study. We then regress the four survey indices on this variable using robust standard errors clustered at the participant level. For all four indices, we find null or positive effects (financial: $\beta = 0.158, p < 0.001$; psychological: $\beta = 0.064, p = 0.118$; cognitive capacity: $\beta = 0.077, p = 0.059$; health: $\beta = 0.054, p = 0.177$), which is inconsistent with the strategic distortion theory and, if anything, suggests that people who think that their answers can affect how much money they get will be more likely to have "good"-sounding outcomes. We also regress each index measure on an interaction of the beliefs dummy with the cash group dummy and find no significant interactions (financial: $\beta = -0.156, p = 0.051$; psychological: $\beta = 0.121, p = 0.138$; cognitive capacity: $\beta = 0.019, p = 0.817$; health: $\beta = 0.053, p = 0.506$).

We conclude that the correlational survey response data are also largely inconsistent with the possibility that cash participants strategically distorted their responses downward to appear needier.

D.1.3 Comparing self-reports against bank data

As discussed in Appendix Section E.2, we find some evidence that relative to the Control group, the cash groups may have been more likely to underreport their savings and income in post-treatment time periods (but not at baseline). Although these data only apply to the sample of people who provided access to their bank account data, we view this result as potentially being consistent with the strategic distortion mechanism.

D.1.4 Discussion

Overall, our experimental and correlational survey response data do not support this pathway, but the comparison of self-reported financial status and administrative bank data does.

D.2 Reference dependence

This section describes the evidence for the reference dependence mechanism: the possibility that cash group participants spent or earmarked their money before taking the t2 survey, compared their lives to a time when they still had money, and subsequently felt worse than if had they never received that money at all. We test this through an embedded experiment and correlational survey responses, both of which were added partway through the trial.

D.2.1 Experiment

The embedded experiment was conducted with Wave 5, 6, and 7 participants at the start of the t3 survey. A total of 2,474 participants were randomly assigned to one of three groups. The AfterPaid group ($N = 798$) was asked to think back to a certain target date, where that date was immediately after the cash participants in that wave had received their money. To make the target date salient, participants were asked to compare their lives on that date to the present day (i.e., the day they were responding to the t3 survey), identifying whether their lives on the target date felt worse, better, or about the same as their lives “today.” The BeforePaid group ($N = 831$) was asked the same question, except the target date was one month before the date used in the AfterPaid group—i.e., about a month before cash group participants in this wave received their money. The NoReferencePoint group ($N = 845$) was not asked any such question.

For instance, cash participants in Wave 7 received their UCT payments on January 28, 2021. On March 4, 2021, they took their t3 survey. Those in the AfterPaid group were asked, “Please think about your life on 1/30/21 and how it compares to your life TODAY. To the best of your memory, would you say that your life on 1/30/21 felt... [Much worse than you feel your life is today, Slightly worse than you feel your life is today, About the same as you feel your life is today, Slightly better than you feel your life is today, Much better than you feel your life is today].” Those in the BeforePaid group were asked the same question, except instead of 1/30/21, they saw 12/30/20.

The reasoning behind this intervention was as follows: If reference dependence played a role in participants’ feelings as described above, then we should see an interaction between this reference

point intervention and being in a cash group. Specifically, for participants in the BeforePaid group who had received cash, the intervention would counteract the natural reference point of thinking about oneself when times were better by focusing participants' attention on a time when they had comparatively little money and times were perhaps worse. Thus, these individuals would feel comparatively better about their lives in the present and report better survey outcomes than cash group participants in the AfterPaid and NoReferencePoint group. On the other hand, for participants in the AfterPaid group who had received cash, the intervention would reinforce the natural reference point of thinking about oneself when times were better, and thus these individuals would report worse or equal outcomes than cash group participants in the BeforePaid and NoReferencePoint groups. Finally, we should observe no difference between the AfterPaid, BeforePaid, and NoReferencePoint groups for participants who had not received cash, as the dates would on average be meaningless to them.

First, though not a central part of the theory test, we examine answers to the intervention question to assess whether the manipulation seems to have shifted participants' likelihood of thinking about "better" and "worse" times. We find that the AfterPaid participants who had also received cash felt that their lives were directionally (though not significantly) better on the target date, relative to cash group participants in the BeforePaid group and relative to Control (no cash) participants in either reference point group (interaction: $\beta = 0.165, p = 0.127$).

In what follows below, we examine the effects of this reference point randomization on t3 index question responses. First, we regress each of the four indices on just binary indicators for the reference point group. We find a negative and significant effect for the AfterPaid group on the financial index ($\beta_{AfterPaid} = -0.110, p = 0.035$) and a negative and marginally significant effect for the BeforePaid group on the same index ($\beta_{BeforePaid} = -0.090, p = 0.077$). We find no main effects on the other indices (all $p \geq 0.141$) in t3. Next, we regress each of the four indices on an interaction of the reference point group and whether the participant had received any cash. We find no significant interactions for either the *AfterPaid* \times *AnyCash* interaction (all $p \geq 0.244$) or the *BeforePaid* \times *AnyCash* interaction (all $p \geq 0.080$).

The results are similar if we regress the subjective financial well-being construct—based on the three subjective response questions that came most immediately after the intervention—on these interactions. They are also similar if we limit the analysis to the only wave for which we ran this manipulation where there were no "special" seasons in the manipulation (the winter holidays and New Year's) that could have interfered with participants' responses (this was Wave 5, $N = 764$). Thus, the experimental evidence is inconsistent with the reference dependence mechanism.

D.2.2 Correlational survey responses

To supplement the experiment, we also use correlational survey responses. In t4, we asked participants two questions intended to measure the extent to which they naturally used reference points that could produce the pattern described above. Participants were asked how much they agreed with two statements: "I frequently compare my life to how it was in better times and then feel bad"

and “When I think about my finances as they are now, I often compare them to times in my past when I was doing worse” (both questions: 1=“Strongly disagree”; 5=“Strongly agree”). A person who agrees with the first statement and disagrees with the second may be more susceptible to the type of theorized reference point comparison. We reverse code answers to the second statement and average across the two statements to create a “negative comparison susceptibility” measure.

We find that this measure negatively predicts all four index outcomes (financial: $\beta = -0.188, p < 0.001$; psychological: $\beta = -0.351, p < 0.001$; cognitive capacity: $\beta = -0.174, p < 0.001$; health: $\beta = -0.354, p < 0.001$), which is consistent with the reference dependence theory. Moreover, we find that the \$2,000 group scores higher on this negative comparison susceptibility measure than the Control group ($\beta = 0.090, p = 0.037$). However, there are no differences between the \$500 group and Control in negative comparison susceptibility ($\beta = 0.029, p = 0.389$), and there are no significant interactions between negative comparison susceptibility and being in a cash group when regressing the four indices on this interaction (all $p \geq 0.115$).

We also test whether this negative comparison susceptibility measure mediates the effect of the treatment on the indices. We regress each of the three indices for which there were negative effects (financial, psychological, and health) first just on the cash group indicators, and then the cash group indicators along with the negative comparison susceptibility measure. The financial index cash group indicators change little in terms of effect size or significance, though one of the psychological index measures does (without the mediator: $\beta_{\$2000} = -0.101, p = 0.015$; with the mediator: $\beta_{\$2000} = -0.058, p = 0.194$), as does one of the health index measures (without the mediator: $\beta_{\$2000} = -0.106, p = 0.010$; with the mediator: $\beta_{\$2000} = -0.069, p = 0.126$). Thus, the correlational survey responses yield somewhat mixed evidence for the reference dependence mechanism.

D.2.3 Discussion

Overall, we conclude that there is some, but not particularly strong, evidence for the reference dependence mechanism in explaining why we fail to observe positive effects of cash on our survey measures.

D.3 Spending behavior

In this section, we test the possibility that UCT recipients spent their money in ways that harmed them, for instance on recreational drugs and alcohol (Banerjee and Duflo, 2007; Brune et al., 2017). We first examine Figure 4 of the main text, which graphs participants’ intended uses of their UCT and survey payments. Both “lotteries and gambling” and “recreational drugs, tobacco, and alcohol” were mentioned very rarely. In fact, the former category was mentioned by only a single participant, and the latter was mentioned by none.

Although responses to this question were non-binding and had no effect on subsequent receipt of funds, it is perfectly plausible that participants nevertheless wanted to obscure their true uses or responded based on what they believed researchers or the non-profit wanted to hear (Godoy,

Karlan and Zinman, 2021). To address this concern, we turn to the administrative financial transactions data, flagging any transaction descriptions that include the words “alcohol,” “liquor,” “beer,” “wine,” “bar,” “tobacco,” “cigarettes,” “cigars,” “smoking,” “cannabis,” “lottery,” “lotto,” “Mega Millions,” “Daily Millions,” “Powerball,” “2By2,” “Keno,” “Cash4Life,” “All or Nothing,” “Caesars,” “casino,” “gambling,” “betting,” “FanDuel,” “DraftKings,” “Betfair,” “BetMGM,” or variants thereof. We find no differences in average daily spending on these items across treatment groups after UCT payment during the trial period ($\beta_{\$500} = \$0.06, p = 0.106; \beta_{\$2000} = \$0.05, p = 0.217$). These results are consistent with prior work failing to find an effect of cash transfers on temptation goods spending (Evans and Popova, 2017; Yoo et al., 2022). Moreover, when we regress post-treatment index outcomes on this “temptation goods” spending measure, we find no relationship (all $p \geq 0.117$). While we can only speak to the participant sample for which we have financial transaction data, and we cannot analyze cash transactions or transactions through unobserved bank accounts, these results suggest that it is unlikely that cash group participants spent substantial amounts of money on these goods and activities, and that this in turn generated worse survey outcomes.

We can also test a somewhat narrower but related hypothesis. Prior work (Gross and Tobacman, 2014) has argued that influxes of cash can increase the likelihood of drug- and alcohol-related emergency room visits. If this were the case in our study, we should see an increase in healthcare expenditures.⁸ But regressing healthcare expenditures over the course of the trial (after UCT receipt) on the treatment group dummies, we find no evidence of this possibility (all $p \geq 0.369$).

Finally, we note that both the smaller and larger UCT amounts seemed to at least weakly increase participants’ patience (Frederick, Loewenstein and O’Donoghue, 2002) ($\beta_{\$500} = 0.27, p = 0.003; \beta_{\$2000} = 0.22, p = 0.058$),⁹ and that patience in turn positively predicts all four index outcomes (examining across all post-treatment time periods; financial: $\beta = 0.040, p < 0.001$; psychological: $\beta = 0.045, p < 0.001$; cognitive capacity: $\beta = 0.031, p < 0.001$; health: $\beta = 0.031, p < 0.001$). These patterns suggest that, although cash can of course increase a person’s access to temptation goods, it may simultaneously decrease a person’s interest in spending on those same goods, which all else equal would predict better survey outcomes. Taken together, these data suggest that it is unlikely that participants spent money in ways that actively harmed them and subsequently suppressed their index outcomes.

D.4 Expectations of money amounts

We next review evidence on the extent to which expectations of how much (more) money one will receive could have contributed to the lack of positive effects of cash. We consider three possible variants.

The first variant proposes that the cash groups were disappointed by the fact that they “only”

⁸Note that, while it may be relatively common for people to purchase drugs or alcohol with cash, health expenditures are more likely to be paid for with cards or checks.

⁹This is consistent with prior work documenting a relationship between financial stress and impatience; see Haushofer and Fehr (2014).

received \$500 or \$2,000, and that this disappointment subsequently pushed their post-treatment survey responses downward. This, however, seems unlikely. As mentioned in Section 2.1, 43% of our participants had received money from the organization prior to the start of this trial, and 99% of those participants had gotten \$500. Participants in our trial may have reasonably set their expectations based on these prior UCT receipts. Because the UCTs in this trial, \$500 and \$2,000, matched or exceeded those initial amounts, we believe it is unlikely that participants (or at least those participants) found them disappointing.

The second variant entertains the idea that cash group participants were disappointed that the UCTs were one-off transfers rather than repeated, and/or that they were experiencing a declining sequence of payments (from \$500 or \$2,000 to only the \$20 survey payments; Loewenstein and Prelec (1993)). However, this would be difficult to reconcile with the fact that we see null and negative treatment effects in t2, just one week after UCT receipt. It seems quite unlikely that participants would have expected to receive multiple UCTs in the span of a week.

The third variant relates to hope on the side of the Control group, as opposed to the disappointment felt by the cash groups. If the Control group was expecting to receive money throughout the course of the entire trial, in theory this hope could have been sufficient to elevate their survey responses in such a way that they looked as good or better than the cash group participants in terms of indices. To test this, we examine responses to a t4 question asking participants for their best guess as to how much more money they would receive from the non-profit, supposing that the non-profit gave them money. The cash group participants reported higher, not lower, estimates ($M_{Control} = \$341$; $M_{\$500} = \504 ; $M_{\$2000} = \811 ; all two-sided t-tests $p < 0.001$). Although by no means a perfect test, this datapoint is somewhat inconsistent with the theory that the Control group participants' hope for more money was sufficient to elevate their responses beyond the cash groups' responses.

We conclude that overall, it is unlikely that expectations of money amounts could explain why we observed no positive effects of cash on the survey index values.

D.5 Inference from money receipt

Participants were informed that the purpose of the study was to examine how different programs affected people who were struggling financially. Thus, it is possible that receiving money from a charity for this study led UCT recipients to infer that they must be poor. This, in turn, could have pushed their survey responses “downward” relative to Control group participants who would not have made this inference.

However, there are several reasons to believe that this mechanism is not responsible for the observed effects. First, the language informing participants of the purpose of the study was used not only in the UCT notifications, but also in the consent form and every survey notification and payment email, meaning that even the Control group participants had seen this same language anywhere between two and nine times, depending on how many surveys they completed and how many emails they read. Thus, although the cash groups were sent this language one additional time

(in the UCT notification email), all groups saw the language multiple times. Second, the notification email also stated that the purpose of the study was to test how different programs affected people who were struggling financially specifically because of the coronavirus pandemic, which prior work has argued is a comparatively less stigmatizing reason for needing financial assistance (Lasky-Fink and Linos, 2022). Finally, the UCT notification email also informed cash group participants that they had been randomly selected to receive the funds. Thus, they should not have updated on their beliefs about themselves in response to receiving cash. And indeed, it appears that at least some participants internalized this information. As shown in Appendix Table H.12, relative to the Control group, cash group participants were equally likely to say that study benefits were determined by participant characteristics/circumstances, and they were more likely to say that benefits were determined randomly. Thus, we believe it is unlikely that this mechanism explains why the cash groups appeared to be doing no better than the Control group on our main survey outcomes.

D.6 Declining social relationships

Here we consider the possibility that receiving cash led to a decline in social relationships between the cash group participants and their friends and family, perhaps due to participants not meeting their friends' and families' expectations about how much money they would share (Portes, 1998; O'Brien, 2012; McNeill and Pierotti, 2021), and/or the participants choosing to distance themselves from their friends and family to avoid being asked for help (Dana, Weber and Kuang, 2007).

We test social tension or strife through two questions. First, we asked participants how their relationships with people outside the household at that time compared with how they usually are (1=“Much worse than usual”; 3=“About average”; 5=“Much better than usual”). We also asked them to what extent they agree with the statement, “My friends and family give me positive energy every day” (1=“Strongly disagree”; 5=“Strongly agree”). We find no differences between treatment groups on either of these questions when examining across all post-treatment surveys (all $p \geq 0.314$). We also find no differences in the former question when just restricting to t2 measurements (both $p \geq 0.230$), the time period in which we might expect to see the strongest pressures to help; no difference in the latter question for the \$2,000 group when just restricting to t2 measurements ($\beta = 0.007, p = 0.916$); and an increase in the \$500 group's sense of friends'/family's positive energy in t2 ($\beta_{\$500} = 0.103, p = 0.029$). If anything, this latter finding is inconsistent with the theory that participants experienced negative social consequences for not providing enough help.

In summary, we do not find any evidence of social tension or a decline in social relationships among the cash groups, relative to Control.

D.7 Attrition

See Section 4.6 of the main text and Appendix Section C for a review of the evidence on the extent to which attrition could have affected the results.

E Main study: Other additional analyses

E.1 Participant expectations

At the end of the t4 survey, all participants were asked, “Please think back to when you signed up for this study. What, if anything, did you EXPECT to receive as part of this study, in addition to surveys and payments you receive for responding to surveys? Please check all that apply.” They were then allowed to choose from six options: money, peer support, information, social worker help, other, and none of the above. We find that, on average, 52% of participants reported that they expected to receive money. There were no differences between the treatment arms in their likelihood of choosing the cash option (logistic regression with binary indicators for each cash group and robust SE : $\beta_{\$500} = 0.007, p = 0.934$; $\beta_{\$2000} = -0.013, p = 0.902$).

E.2 Reported vs. administrative savings and income figures

For the subset of people for whom we have access to both survey and bank data, we can compare answers to two survey questions—savings and income—against administrative bank data.

E.2.1 Methods

The savings question asked about the participant’s total household savings, including cash, bank account balances, stock values, retirement savings, and investment values (e.g., in a home or car). We compare this to a participant’s current bank balance summed across all observed accounts on the day the survey was sent.

The income survey measure is calculated from two questions: household earned income over the past month (including wages, salary, commissions, bonuses, and tips) and household unearned income over the past month (including public assistance or welfare payments, stimulus checks, tax refunds, payments from non-profit organizations, and unemployment benefits). We use the sum of these two answers to identify “total reported household income.” We compare this to total inflows across all observed bank accounts over the 30 days before the survey was sent. If data are observed only for part of the 30-day period, we use the partial data, taking the sum of all inflows over the observed days.

Because the savings and income questions were asked in each of the four surveys, and in most cases we observe bank data for at least a couple months, we can have up to four “reported vs. administrative” savings datapoints and four “reported vs. administrative” income datapoints per person.

It is important to note that the survey and administrative bank measures are capturing slightly different constructs for a few reasons, and it is not clear which value we would typically expect to be higher. First, the survey questions asked about household savings and income, whereas the accounts we observe may be held by individuals, single households, or multiple households. Because we unfortunately do not observe bank account ownership, the effect of this point on the direction

of the gap between reported and administrative figures is ambiguous. Second, some inflows we observe in the bank accounts may not have been “new” income, instead, for instance, being a transfer from a different unobserved account, a deposit from previously stored cash, or a friend or relative repaying a loan. Because participants were not asked to include such inflows in their survey responses, this point would tend to push reported income to be lower than administratively-observed income. Finally, the survey questions asked about *total* savings (including cash, the value of physical assets, and savings in bank accounts other than the one(s) we observe) and *total* income (including income received as cash and never deposited or deposited into bank accounts other than the one(s) we observe). Because participants were asked to report money that in theory we could have administratively observed, but also money we could not have observed, this point would push reported savings and income to be higher than their administrative counterparts.

In other words, even if participants perfectly reported their household’s savings and income, the survey questions and corresponding administrative data measures would likely yield different numbers. Moreover, the net direction of the bias is unclear. Nevertheless, we compare the figures we have to identify how (dis)similar they may be from one another.

E.2.2 Results

The dataset for these analyses necessarily only includes people who provided a self-report of savings (or income) at a given survey time period *and* for whom we observe the corresponding bank data for that same period. We observe at least one “reported vs. administrative” savings datapoint for 1,963 participants (37% of the sample; 4,446 total datapoints) and at least one “reported vs. administrative” income datapoint for 2,099 participants (40% of the sample; 5,534 datapoints).

Firstly, and most importantly, we find that across all time periods, there is a positive and statistically significant correlation between reported savings and observed bank balances ($r = 0.386, p < 0.001$), and between reported income and observed account inflows ($r = 0.225, p < 0.001$). The positive correlation also holds when examining individual time periods.

To identify any gaps between the reported savings (or income) figures and what we observe in the bank data, we first subtract the administrative figures from the self-reports (winsorizing at 90%, top and bottom), then plot the distribution of differences. See Appendix Figure G.11. We find there is slightly more mass to the left of 0, which could reflect (i) the fact that the survey questions and bank measures captured slightly different constructs, and/or (ii) potential underreporting (reporting that one has less savings than one has, or reporting that one has received less income than one did). Across all time periods, the gap between reported and administrative savings is quite small (median: $-\$0.27$; average: $-\$61.02$, two-sided t-test testing for difference from 0: $p < 0.001$). For income, the gap is larger (median: $-\$124.13$; average: $-\$271.45$, two-sided t-test testing for difference from 0: $p < 0.001$). To the extent that this is evidence of underreporting, this would be consistent with prior work (Moore, Stinson and Welniak, 2000).

Finally, there is some evidence that the gap between reported and administrative measures is larger among the cash groups, relative to Control, though the differences are not always statistically

significant. On the extensive margin, we observe that in t1, there are no differences between the Control and cash groups in how frequently they report less than is administratively observed for either savings (Control average: 55%; cash average: 55%; two-sided t-test: $p = 0.872$) or income (Control average: 56%; cash average: 55%; two-sided t-test: $p = 0.458$). To measure the frequency of reporting less than is administratively observed in t2 through t4, we calculate the fraction of instances per person where they report less than is administratively observed. For savings, there are no statistically significant differences between Control and cash (Control average: 59%; cash average: 61%; two-sided t-test: $p = 0.291$). For income, the frequency is higher among the cash groups than Control (Control average: 65%; cash average: 71%; two-sided t-test: $p = 0.001$).

On the intensive margin, again examining the distribution of differences between reported and administrative savings and income, we see no differences between the Control and cash groups in t1 for either savings (OLS regression with robust *SEs*: intercept=−\$57.01, $\beta = -\$10.31, p = 0.525$) or income (intercept=−\$342.21, $\beta = \$156.96, p = 0.060$). In the post-treatment time periods, however, we observe a larger gap among the cash groups for both (OLS regression with robust *SEs* clustering at the participant level to account for multiple observations per person: savings: intercept=−\$151.73, $\beta = -\$114.44, p < 0.001$; income: intercept=−\$1379.78, $\beta = -\$588.43, p < 0.001$). To the extent that this is evidence of strategic underreporting to get more benefits (see Appendix Section D.1), this would be consistent with prior work (Martinelli and Parker, 2009).

E.3 Measuring treatment details

We use two metrics to assess whether the participants were treated. The first metric asks: did the organization send participants money according to their treatment group assignment? Using the non-profit organization’s payment data, we calculate how much money each participant was sent during the trial (i.e., from the first day that the t1 survey was open to the last day the t4 survey was open). This sum thus excludes the completion bonuses and lotteries, but includes the UCT payments, survey payments, and bank-account-linking payments. We find that the amounts sent align with treatment group assignment: the average amounts were \$65 for the Control group, \$572 for the \$500 group, and \$2,056 for the \$2,000 group.

We tag a participant as having been sent the correct amount if they were sent the UCT amount they were randomized to plus up to \$100 during the trial period (this allows for four survey payments at \$20 apiece plus up to two financial account linkings at \$10 apiece). Using this metric, we find that 98% of the Control group was sent the correct amount (with the remaining participants all being paid “too much”), 96% of the \$500 group was sent the correct amount (with an additional 2% being paid “too much”), and 96% of the \$2,000 group was sent the correct amount (with an additional 3% being paid “too much”).

The second metric captures how much money participants received in their external accounts (bank accounts or prepaid cards) during the trial. We find that participants received most, though not all, of the money that was sent to them: the Control group on average received \$56, the \$500 group received \$515, and the \$2,000 group received \$1,792. These sums correspond to 70%, 89%,

and 87% of what each person in those treatment groups was sent, respectively.

E.4 Effect of UCTs on transaction volume, debit volume, and net expenditures

Before the UCT payment date, the average number of daily transactions across treatment groups was 1.76 and the average number of daily debits was 1.35, with the \$500 group but not the \$2,000 group having somewhat higher activity (transactions: $\beta_{\$500} = 0.27, p = 0.005; \beta_{\$2000} = 0.11, p = 0.332$; debits: $\beta_{\$500} = 0.22, p = 0.004; \beta_{\$2000} = 0.07, p = 0.473$). This amounted to an average daily expenditure of \$64.67, again with the \$500 group but not the \$2,000 group spending more than Control ($\beta_{\$500} = \$9.00, p = 0.005; \beta_{\$2000} = \$3.07, p = 0.415$). The net expenditures (debits minus credits) amounted to an average of \$1.77 per day, with no differences by treatment group ($\beta_{\$500} = \$0.59, p = 0.228; \beta_{\$2000} = -\$0.45, p = 0.457$).

In the first two weeks following the UCT, we observe increased economic activity for both cash groups using three different measures. First, both cash groups had a higher number of daily transactions than Control ($\beta_{\$500} = 0.50, p < 0.001; \beta_{\$2000} = 0.83, p < 0.001$). Second, they had a higher number of daily debits relative to Control ($\beta_{\$500} = 0.43, p < 0.001; \beta_{\$2000} = 0.71, p < 0.001$). Finally, while the Control group had average daily net expenditures of \$0.63 in the two weeks following the UCT payment date, the \$500 group had daily net expenditures of \$20.58 ($p < 0.001$) and the \$2,000 group had daily net expenditures of \$72.63 ($p < 0.001$). For all three measures, however, the effects quickly dissipate, with the groups being largely statistically indistinguishable by the second fortnight. For all three measures, results are similar when including regression controls.

E.5 Benchmarking survey result effect sizes

Here, we benchmark the effect of having received a UCT on a survey index against the effect of having lost one's job because of the pandemic on an index. We use this job loss measure because it is a plausibly exogenous negative shock that affected about half (49.6%) of participants, and because data for this variable is available for all participants (having been measured in the baseline survey). Participants could have lost their job up to 14 months before responding to one of the post-treatment surveys.

To estimate the relative effects of having received a UCT versus having lost one's job, we first regress each of the four indices on a binary indicator for having received any UCT, examining post-treatment time periods together and clustering robust *SEs* at the participant level. We then run all four regressions again, replacing the treatment group assignment indicator with the binary indicator for having lost one's job as a result of the pandemic. Finally, we divide the UCT regression β by the job loss regression β for each index. We find that for the financial index, receiving a UCT corresponds to 0.22 of the effect of losing one's job; for the psychological index, it is 0.28; for the cognitive capacity index, it is 1.07; and for the health index, it is 0.67.

E.6 Objective and subjective outcome analysis details

In this section, we provide more details on the effects of cash on objective and subjective survey measure responses (see Section 4.3 of the main text). Two coders categorized all valenced survey questions as either being “more objective” or “more subjective.” The objective questions were savings, earned income, debt, employment, whether they were housed (defined as not staying with friends or family, not in a shelter, and not experiencing homelessness), whether they had housing debt, the interest rate they would pay on a \$500 loan, the Raven’s matrices score, nutrition, and exercise. The subjective questions were the ones remaining: work performance and satisfaction; liquidity constraints; the three subjective financial well-being questions; the extent to which the participant felt they could meet their household’s needs and wants; how stressed they felt about how to spend their money; whether they felt they could pay for everything over the next 30 days and if not, how much more they would need; the two agency questions; the extent to which the participant felt as though they were living their best life; the three positive mental health questions; happiness; anxiety; loneliness; the eight depression questions; how overwhelmed they felt by others’ needs; how much support the participant had from their community; perceived parenting quality; their perceived relationship(s) with their child(ren), partner, and others outside the household; the extent to which others outside the household gave them positive energy; the two “money on the mind” questions presented in each time period; the four self-assessed memory questions; self-assessed general physical health; sleep quality; and food security.

We conduct two analyses to test whether the treatment had differential effects on more objective versus subjective survey outcomes. In the first analysis, which focuses on effect sizes, we begin by orienting all variables such that higher values are better and Z-scoring each variable separately, thereby putting all the variables on a common, comparable scale. (We construct Z-scores by taking the variable value at a given time period, subtracting the mean of that variable across all treatment arms at t1, and dividing by the SD across all treatment arms at t1. For the Raven’s matrices and “needs 30 days” variable, we do not have a t1 measure, and thus we standardize using the mean and SD of the Control group at each time period.) We then take the average of all the objective Z-scored variables, and separately take the average of all the subjective Z-scored variables. Subsequently, we perform two regressions. The first involves the average Z-scored objective variables, which are regressed against treatment group indicators. We aggregate across all post-treatment values and employ robust standard errors clustered at the participant level. Subsequently, we repeat this process for the average Z-scored subjective variables.

We find that the coefficients in the objective variable index regression are negative but not significantly different from Control ($\beta_{\$500} = -0.021, p = 0.189$; $\beta_{\$2000} = -0.029, p = 0.136$). On the other hand, the coefficients in the subjective variable index regression are more negative and statistically significantly different from Control ($\beta_{\$500} = -0.050, p = 0.010$; $\beta_{\$2000} = -0.082, p = 0.001$).¹⁰ These results hint at the possibility that, to the extent that we observe negative outcomes

¹⁰In robustness checks, we also construct indices using the Anderson (2008) approach. We find a similar, albeit somewhat weaker, pattern: for the objective index, neither treatment group differs from Control ($\beta_{\$500} = -0.027, p =$

for three of the four survey indices in our main analyses, the effects stem more from the subjective, rather than objective, outcome reports.

One potential concern is that the observed null on the objective question index stems from the fact that the objective variables were measured less precisely. Thus, as a second analysis, we examine each variable individually instead of using an average for the two categories. Specifically, we regress each of the objective and subjective survey variables on binary indicators for each treatment group, examining all post-treatment measures together and using robust *SEs* clustered at the participant level. We use logistic regressions for the two binary dependent variables (employment and whether one has housing debt) and OLS otherwise. For each regression, we count the number of treatment group dummy variable coefficients that indicate that the cash group is doing worse than the Control group and is statistically significant at the $\alpha = 0.05$ level.

For the objective variables, the denominator is 20 (2 treatment group coefficients \times 10 variables). Of these, 4 (20%) meet our criteria of indicating that cash group participants scored worse than the Control group participants and were significant at the 0.05 level.¹¹ For the subjective variables, the denominator is 86 (2 treatment group coefficients \times 43 variables). Here we see a much larger fraction of the coefficients meeting our criteria: 44 (51%) indicate that cash group participants scored worse than the Control group participants at the 0.05 level. A two-sample test of proportions indicates that this difference is statistically significant ($p = 0.012$). The results are qualitatively similar if instead of examining each question separately, we use constructs (e.g., we take the mean of the three positive mental health questions and treat it as one dependent variable instead of three) ($p = 0.010$). See Figure 8 in the main text for a simplified depiction of this analysis. We conclude that we cannot reject the null that cash had no effect on responses to more objective questions, but it may have had a negative effect on responses to more subjective questions.

E.7 Wave 6 narrower within-subjects analysis

The administration error in Wave 6 allows us to conduct a within-subjects analysis on a shorter time frame than is typically possible with the design (that is, shorter than the one month that elapses between t1 and t2), examining each participant's change in their survey responses after receiving (or not receiving) a UCT. We limit the sample to just Wave 6 participants who answered both the t2 surveys and answered the first t2 survey before the UCT payment was administered. The sample includes 680 participants, each with two observations. The average number of days between these two surveys is 10.1, and the median is 9. For each of the indices, we take the difference between the two t2 measurements, then run the standard regressions described in Section 4.1 of the main text.

We find no evidence that either cash amount changed participants' index outcomes (five of eight

$0.392; \beta_{\$2000} = -0.039, p = 0.321$), while for the subjective index, the \$500 group is more negative and marginally significantly different from Control ($\beta_{\$500} = -0.048, p = 0.071$) and the \$2,000 group is similarly more negative and statistically significantly different from Control ($\beta_{\$2000} = -0.103, p = 0.004$).

¹¹There is a weak positive effect of cash on the Raven's matrix scores, though this could be due to reciprocity-driven effort rather than true cognitive ability, per se.

β coefficients are negative and none are statistically distinguishable from 0; all $p \geq 0.065$). This conclusion also holds in specifications controlling for the number of days between the t2 surveys and a range of demographics and financial variables (all $p \geq 0.112$). These results suggest that in this narrower within-subjects analysis, which could provide some extra statistical power from reduced noise (though counterbalanced by the reduction in power from a smaller N), the cash still had no detectable positive effect on survey outcomes.

F Prediction study: Additional figures

Predicted Effect Sizes - Indices & Individual Questions

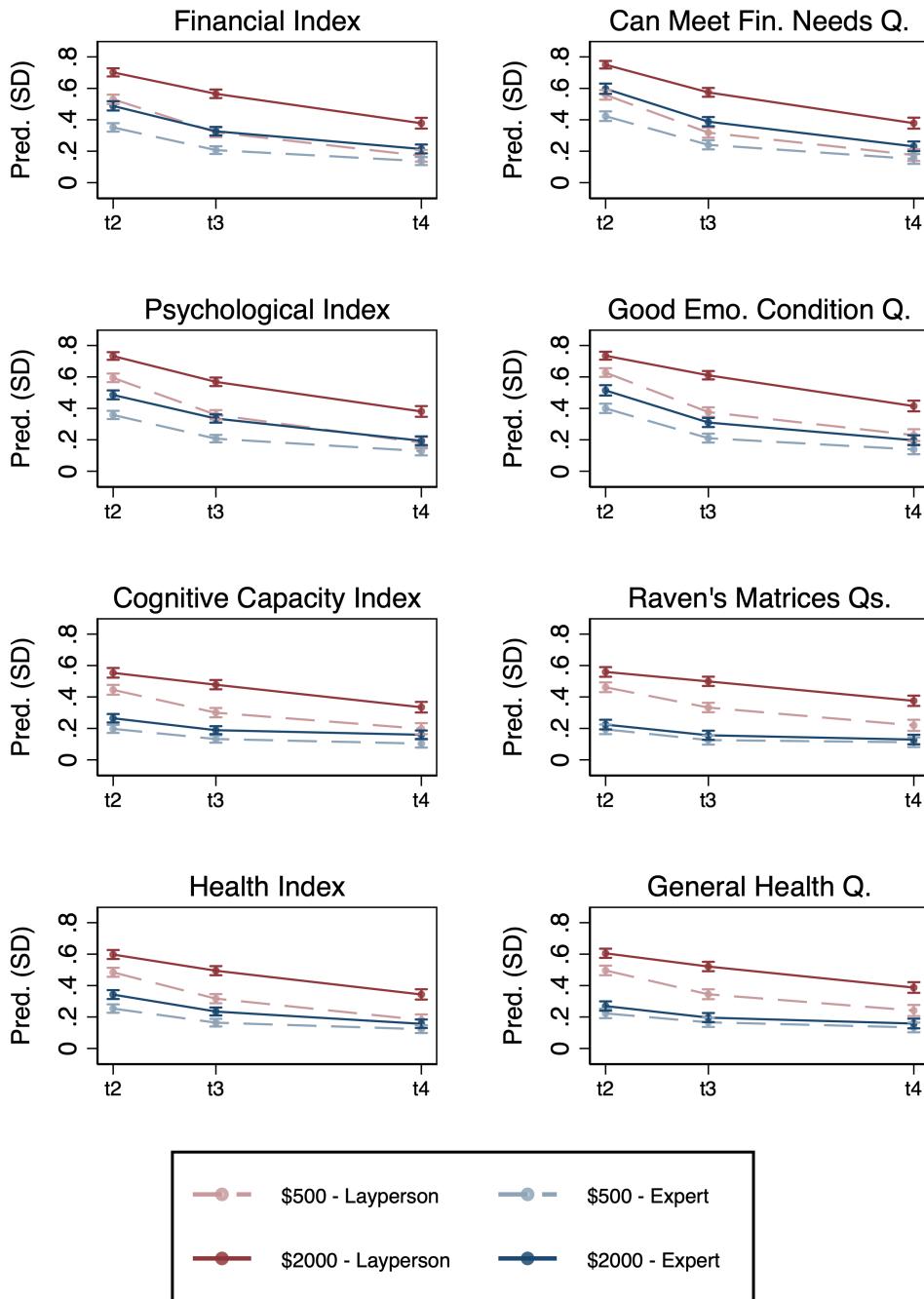


Figure F.1: Experts' and laypeoples' predicted effect sizes (in SDs) for the two treatment groups relative to Control for each index (left) and four survey questions representative of those indices (right). See Appendix Section A for individual question descriptions. Error bars denote 95% CIs.

Predicted vs. Actual Effect Sizes

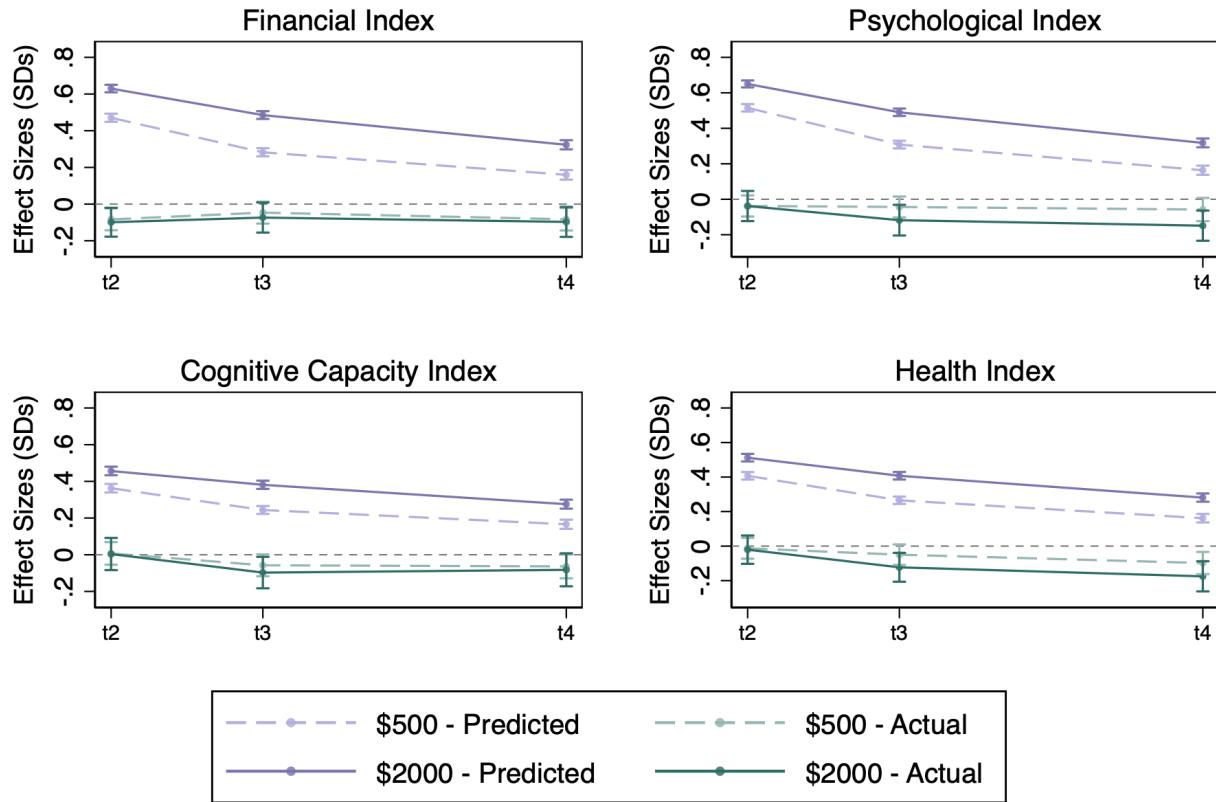


Figure F.2: Predicted and actual effect sizes for the two cash treatments relative to Control at each post-treatment time period. Predictions use both experts' and laypeople's responses. Individual panels correspond to the four survey indices. Error bars denote 95% CIs.

G Main study: Additional figures

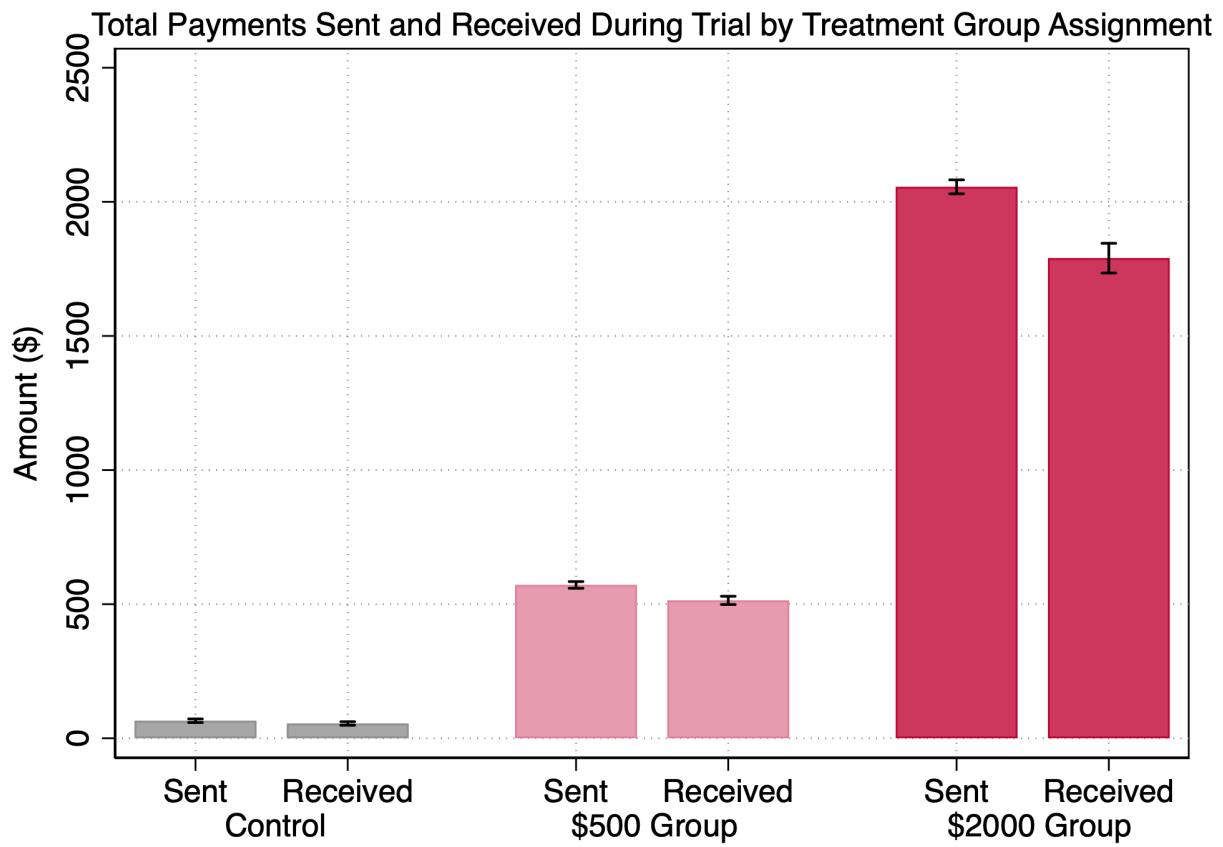


Figure G.3: Administrative data from the partner non-profit organization: all financial payments (including UCTs and survey payments) sent and received between the day the first survey was sent and the closing day of the last survey. “Sent” payments are those that the non-profit organization deposited into the participant’s online platform account or external account (e.g., bank account or payment card). “Received” payments are those that ultimately arrived in the participants’ external accounts. Error bars denote 95% CIs.

Financial Index Variables - Standardized Outcomes

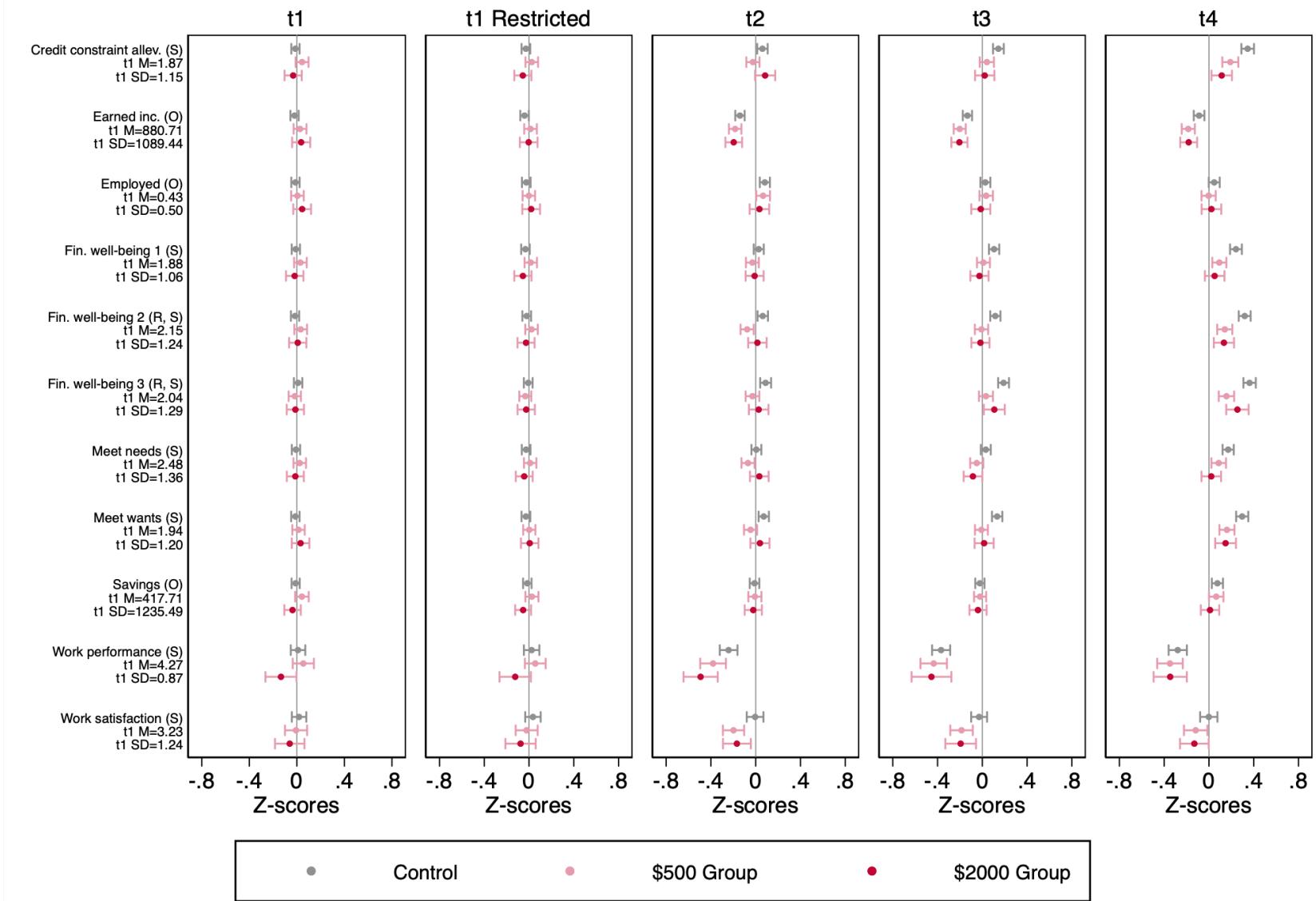


Figure G.4: Variable labels: Unstandardized (raw) mean and *SD* at t1 across treatment arms. See survey instrument for question text and scales. “(S)” and “(O)” = more subjectively (S) or objectively (O) measured. Panels “t1,” “t2,” “t3,” and “t4”: Z-scores (calculated using all participants’ mean and *SD* at t1) and 95% CIs. Panel “t1 Restricted”: Same as panel “t1,” except restricted only to participants who subsequently responded to at least one post-treatment survey. Z-scores are still calculated using all participants’ t1 values, regardless of their subsequent responsiveness.

Psychological Index Variables - Standardized Outcomes

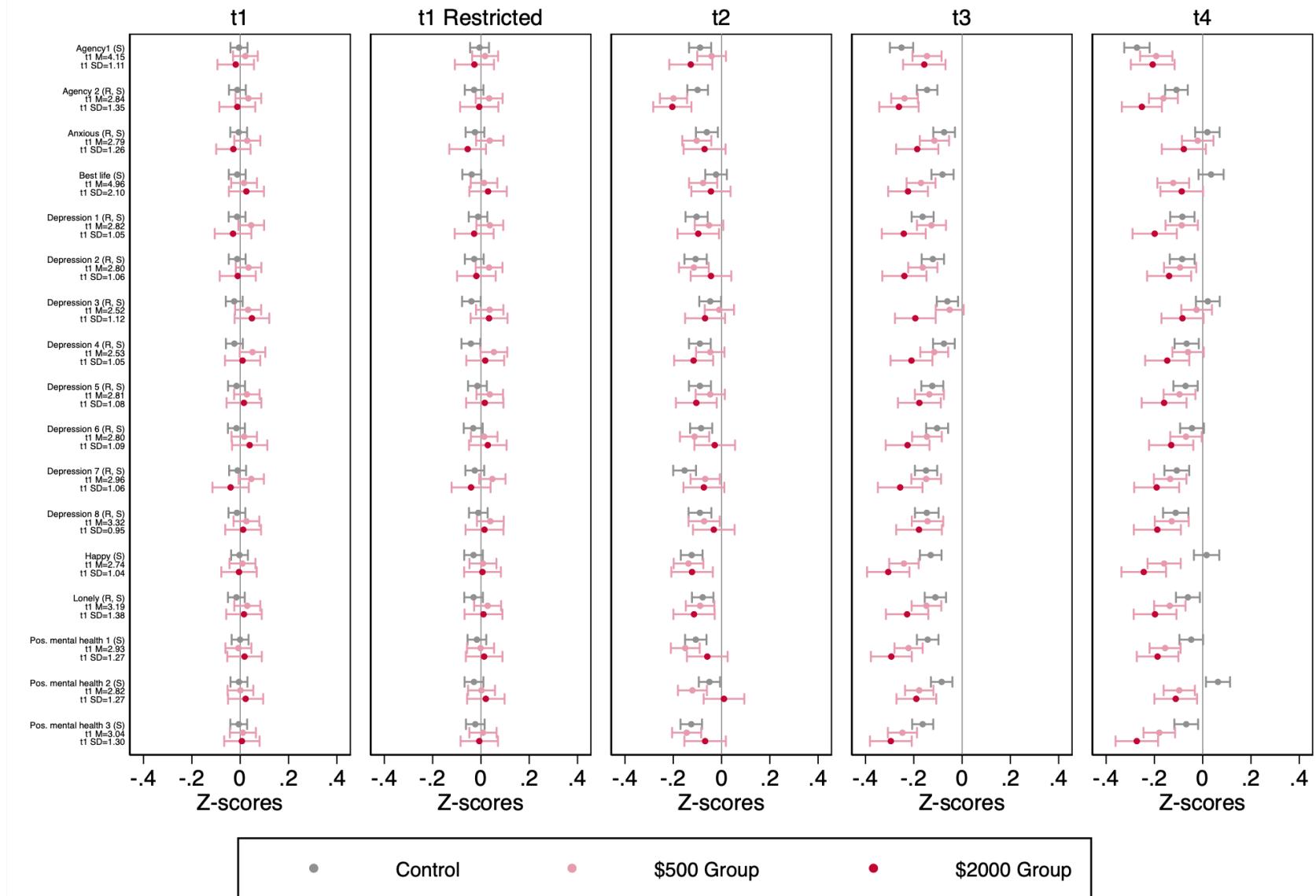


Figure G.5: Variable labels: Unstandardized (raw) mean and SD at t1 across treatment arms. See survey instrument for question text and scales. “(S)” and “(O)” = more subjectively (S) or objectively (O) measured. Panels “t1,” “t2,” “t3,” and “t4”: Z-scores (calculated using all participants’ mean and SD at t1) and 95% CIs. Panel “t1 Restricted”: Same as panel “t1,” except restricted only to participants who subsequently responded to at least one post-treatment survey. Z-scores are still calculated using all participants’ t1 values, regardless of their subsequent responsiveness.

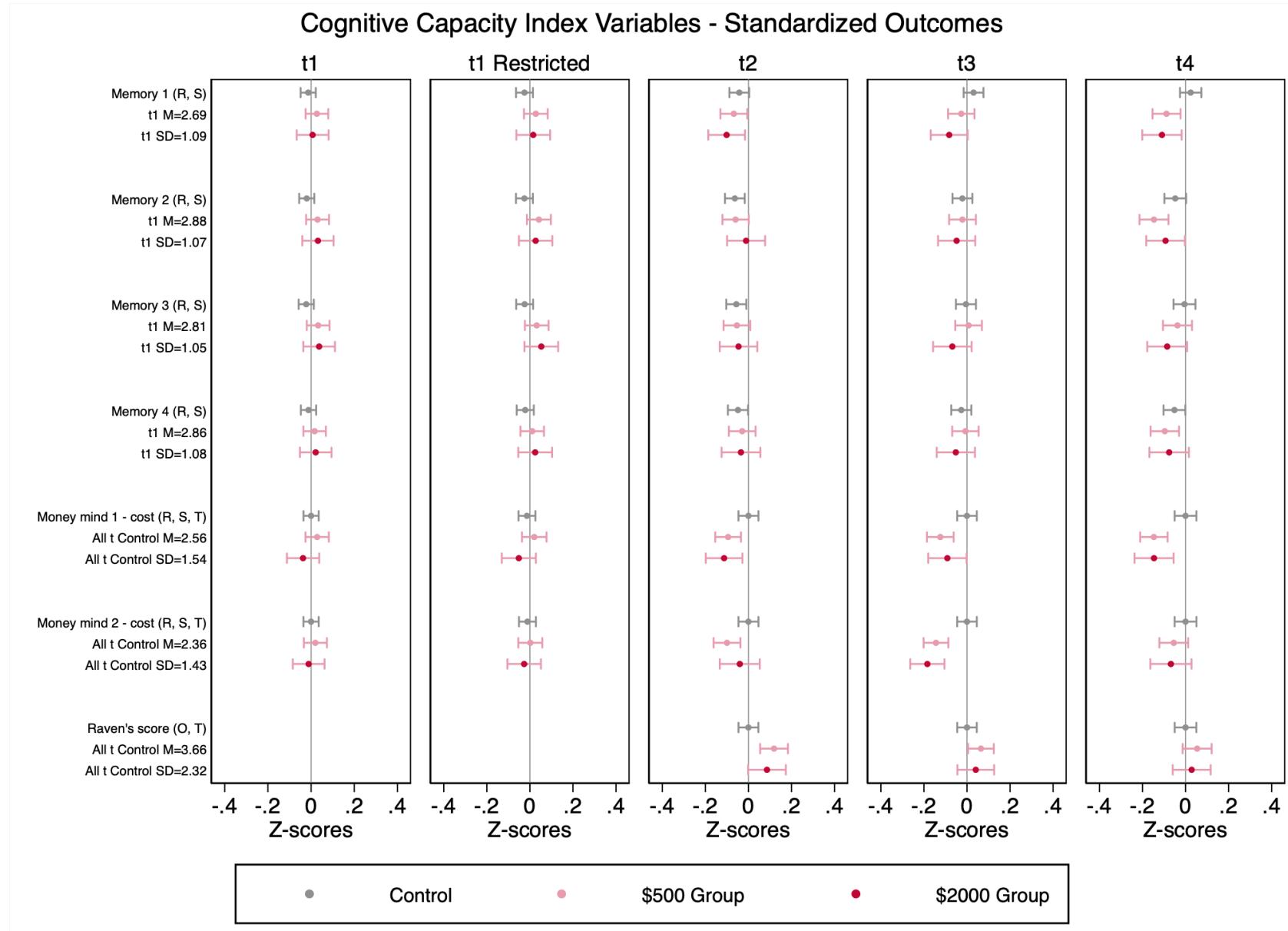


Figure G.6: Variable labels: Unstandardized (raw) mean and *SD* at t1 across treatment arms. See survey instrument for question text and scales. “(T)” = not measured in t1 / questions changed across t’s; values reflect Control group mean and *SD* across all time period(s) measured. “(R)” = reverse coded, s.t. higher values are better. “(S)” and “(O)” = more subjectively (S) or objectively (O) measured. Panels “t1,” “t2,” “t3,” and “t4”: Z-scores (calculated using all participants’ mean and *SD* at t1) and 95% CIs. “(T)” = Z-scores calculated using Control group’s mean and *SD* in that time period. Panel “t1 Restricted”: Same as panel “t1,” except restricted only to participants who subsequently responded to at least one post-treatment survey. Z-scores are still calculated using all participants’ t1 values, regardless of their subsequent responsiveness.

Health Index Variables - Standardized Outcomes

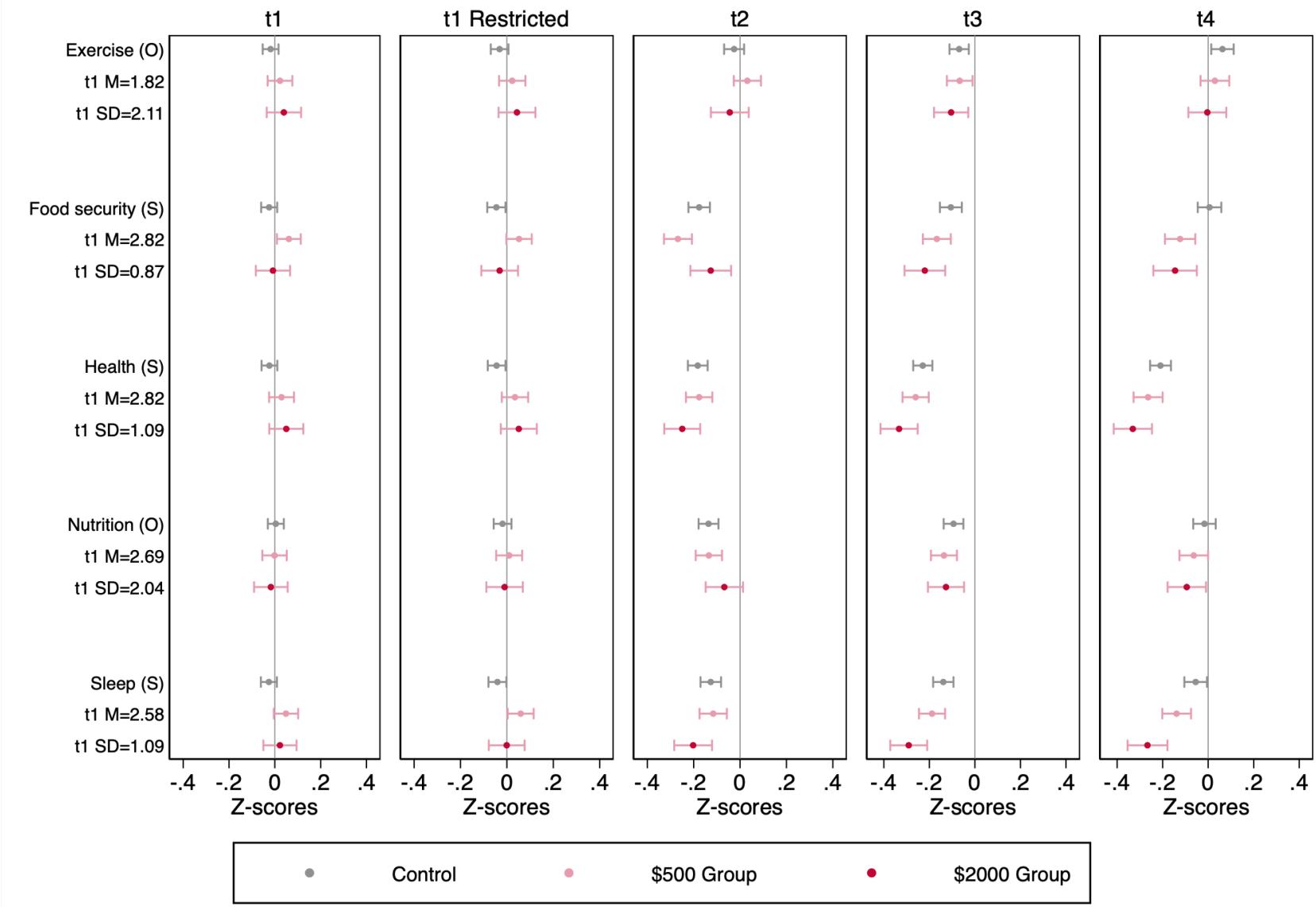


Figure G.7: Variable labels: Unstandardized (raw) mean and *SD* at t1 across treatment arms. See survey instrument for question text and scales. “(R)” = reverse coded, s.t. higher values are better. “(S)” and “(O)” = more subjectively (S) or objectively (O) measured. Panels “t1,” “t2,” “t3,” and “t4”: Z-scores (calculated using all participants’ mean and *SD* at t1) and 95% CIs. Panel “t1 Restricted”: Same as panel “t1,” except restricted only to participants who subsequently responded to at least one post-treatment survey. Z-scores are still calculated using all participants’ t1 values, regardless of their subsequent responsiveness.

Other Valenced Variables - Standardized Outcomes

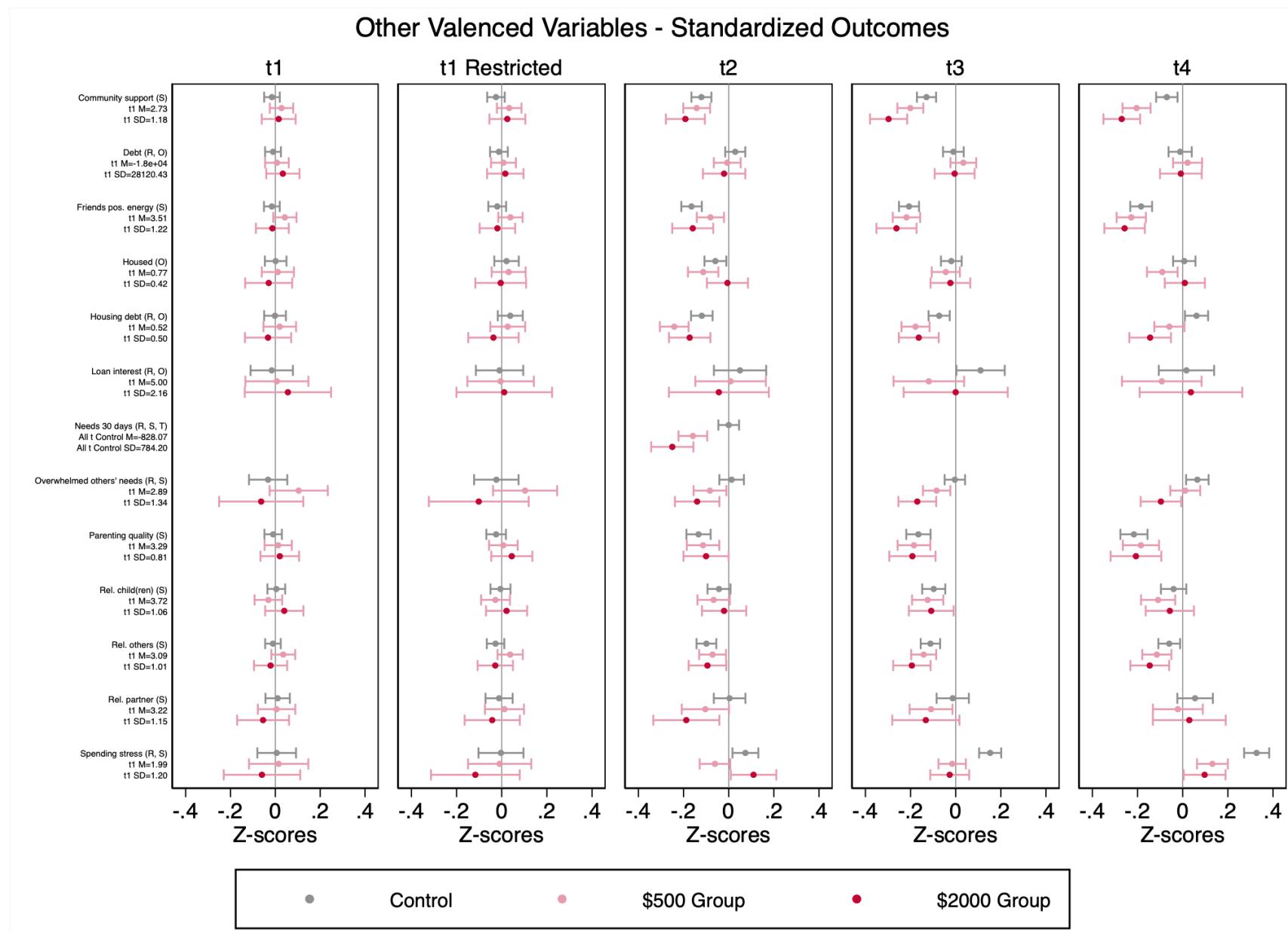


Figure G.8: Outcomes for inherently valenced variables. Variable labels: Unstandardized (raw) mean and *SD* at t1 across treatment arms. See survey instrument for question text and scales. “(T)” = not measured in t1; values reflect Control group mean and *SD* across all time period(s) measured. “(R)” = reverse coded, s.t. higher values are better. “(S)” and “(O)” = more subjectively (S) or objectively (O) measured. Panels “t1,” “t2,” “t3,” and “t4”: Z-scores (calculated using all participants’ mean and *SD* at t1) and 95% CIs. “(T)” = Z-scores calculated using Control group’s mean and *SD* in that time period. Panel “t1 Restricted”: Same as panel “t1,” except restricted only to participants who subsequently responded to at least one post-treatment survey. Z-scores are still calculated using all participants’ t1 values, regardless of their subsequent responsiveness.

Non-Valenced Variables - Standardized Outcomes

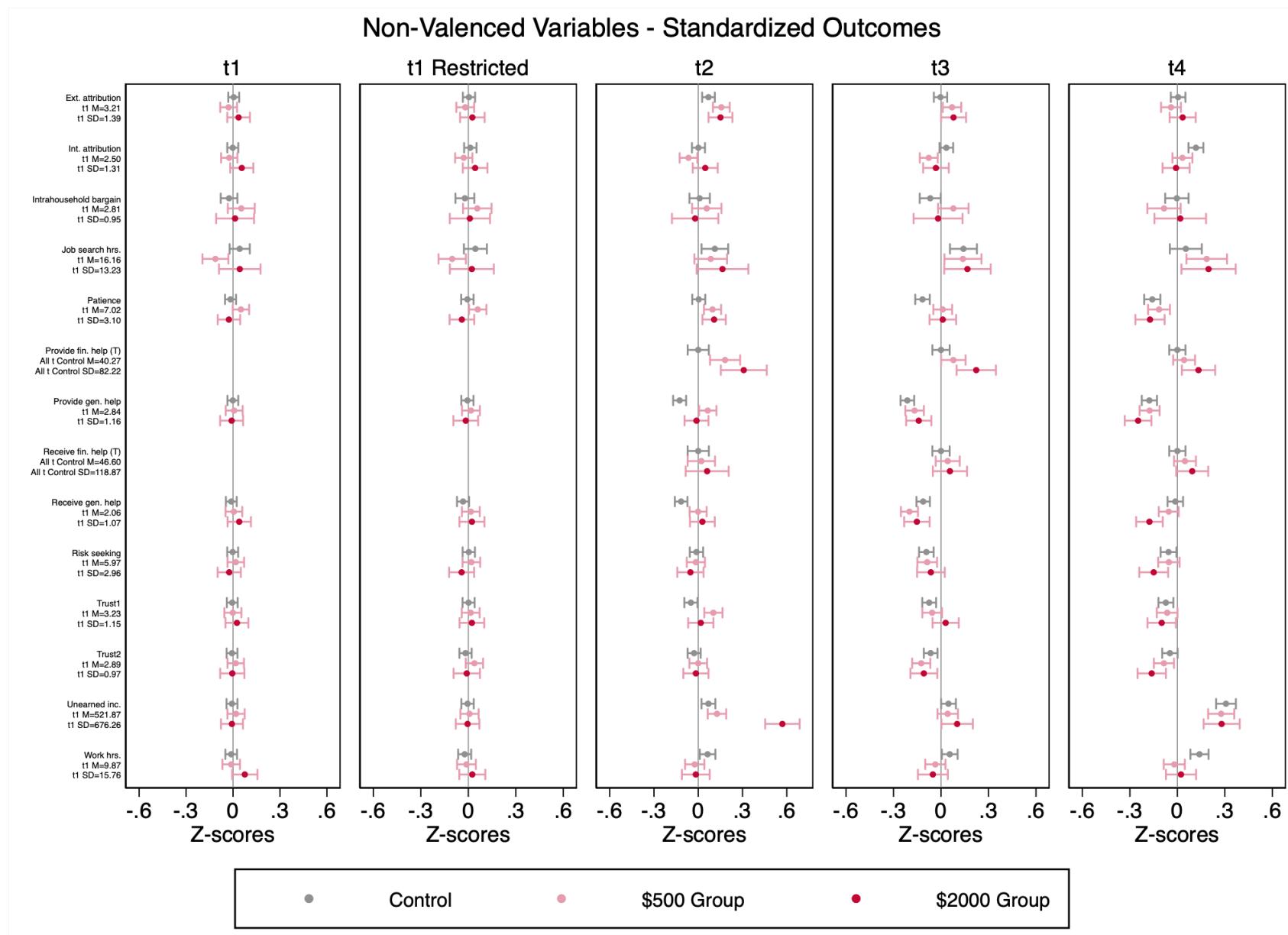


Figure G.9: Outcomes for non-valenced variables. Variable labels: Unstandardized (raw) mean and *SD* at t1 across treatment arms. See survey instrument for question text and scales. “(T)” = not measured in t1; values reflect Control group mean and *SD* across all time period(s) measured. “(R)” = reverse coded, s.t. higher values are better. Panels “t1,” “t2,” “t3,” and “t4”: Z-scores (calculated using all participants’ mean and *SD* at t1) and 95% CIs. “(T)” = Z-scores calculated using Control group’s mean and *SD* in that time period. Panel “t1 Restricted”: Same as panel “t1,” except restricted only to participants who subsequently responded to at least one post-treatment survey. Z-scores are still calculated using all participants’ t1 values, regardless of their subsequent responsiveness.

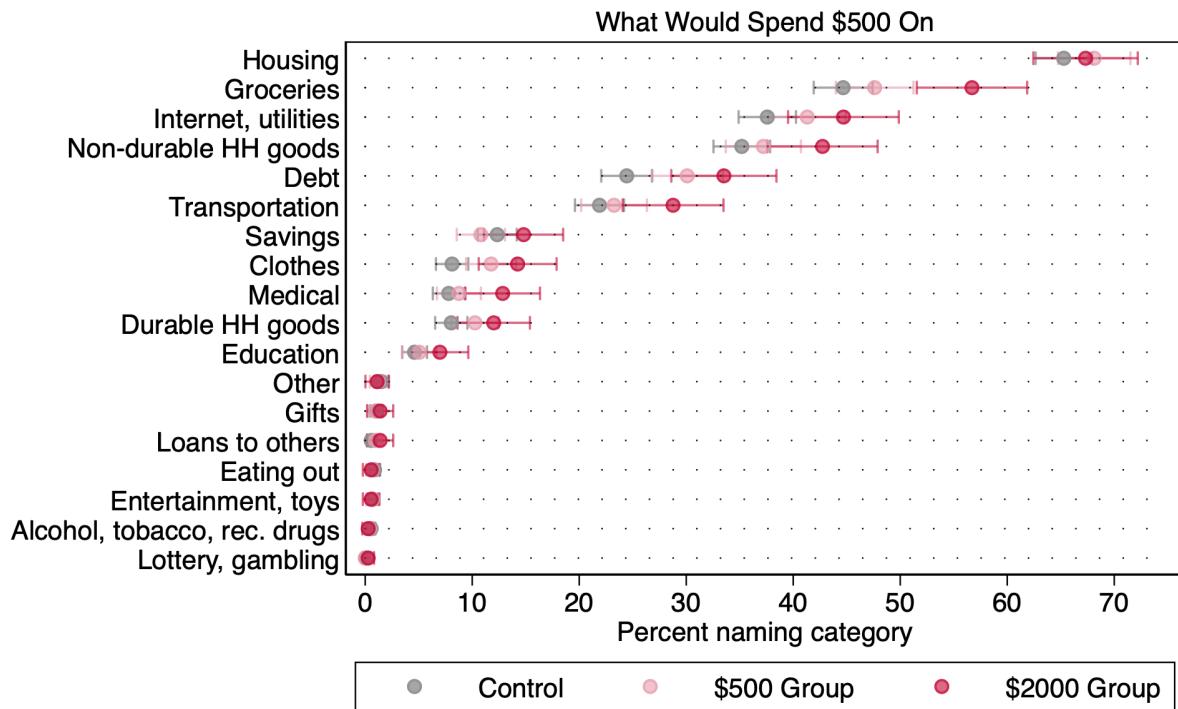


Figure G.10: Responses to the question, “Imagine that the government decided to give everyone a \$500 stimulus check. If you got this money today, what are the MAIN thing(s) you would spend the money on?” HH=Household, rec. drugs=recreational drugs. Circle represents mean value, bars denote 95% CIs.

Distributions of Differences Between Reported and Administrative Bank Data

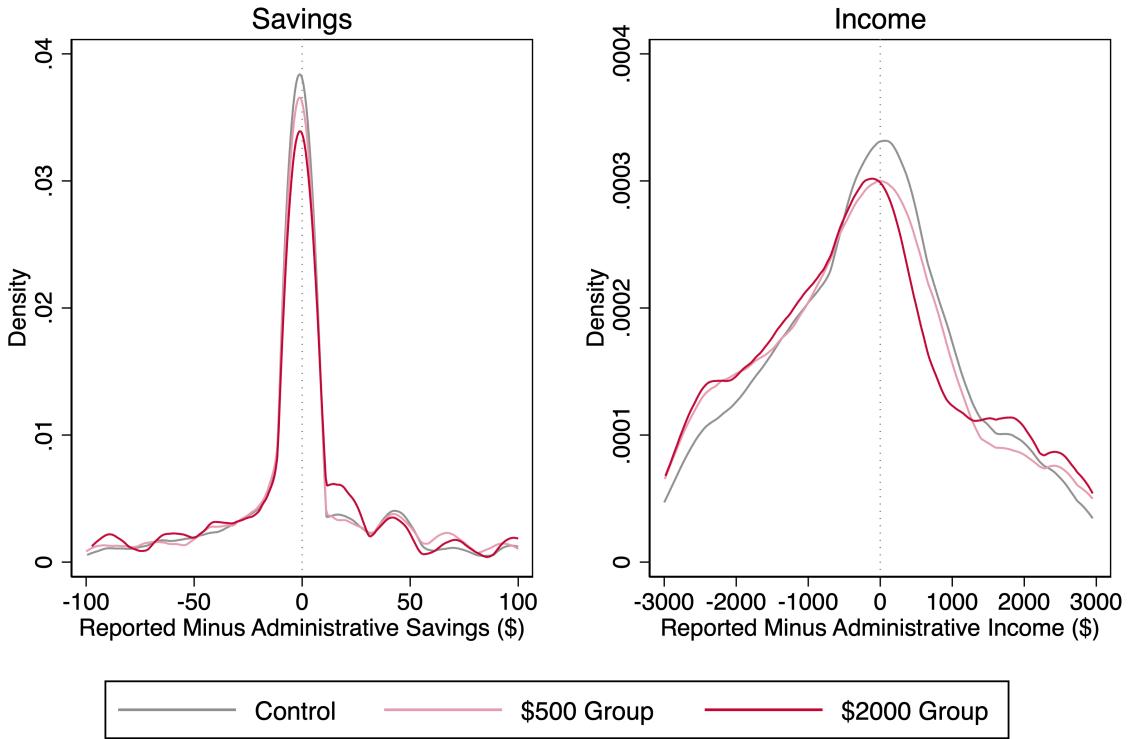


Figure G.11: Epanechnikov kernel density functions. Left panel: Reported total household savings (all cash, bank balances, investment values, etc.) at a given survey time minus observed current bank balance on the day the corresponding survey was sent. Bin width: \$5. Right panel: Sum of reported total household income (earned plus unearned) over the past month minus observed bank inflows over the same month. Bin width: \$300. Note that the survey questions do not exactly mirror available bank data; see Appendix Section E.2. Differences winsorized at the 90% level, top and bottom.

H Main study: Additional tables

Enrollment and Survey Responsiveness.

Wave	t1 Date	Group	Survey					Total Number of Responses	Post-Treatment Survey Response Rate
			t1	t2	t2b	t3	t4		
1	7/24/20	Control	59	25	0	49	36	169	0.90
		\$500	20	8	0	19	15	62	1.00
		\$2000	14	9	0	12	12	47	0.93
2	8/7/20	Control	21	13	0	17	13	64	0.95
		\$500	8	4	0	6	6	24	0.88
		\$2000	4	4	0	3	3	14	1.00
3	10/24/20	Control	132	95	0	97	56	380	0.84
		\$500	52	47	0	46	30	175	0.94
		\$2000	25	24	0	23	20	92	1.00
4	11/5/20	Control	654	495	0	458	390	1997	0.87
		\$500	279	233	0	227	205	944	0.94
		\$2000	141	110	0	118	109	478	0.91
5	12/1/20	Control	710	509	0	449	409	2077	0.82
		\$500	313	267	0	256	233	1069	0.93
		\$2000	159	135	0	125	126	545	0.91
6	12/17/20	Control	1164	680	588	697	561	3690	0.75
		\$500	518	373	349	396	350	1986	0.87
		\$2000	274	180	174	198	187	1013	0.86
7	1/4/21	Control	430	224	0	241	207	1102	0.73
		\$500	184	130	0	133	124	571	0.86
		\$2000	82	57	0	56	53	248	0.77

Table H.1: Survey t1 was the baseline survey. Survey t2b was administered only to Wave 6 and is used in lieu of the first t2 survey for that wave in most analyses. Non-response to specific surveys (e.g., t2) does not imply non-participation in subsequent surveys (e.g., t3, t4). Each survey is independent, and participants may respond to later surveys even if an earlier survey was missed. The last column indicates the proportion of participants who completed at least one post-treatment survey.

Predictors of Providing Bank Data.

	(1)	(2)	(3)
\$500 Group	0.457 (0.065)	0.279 (0.072)	0.296 (0.075)
\$2000 Group	0.553 (0.084)	0.388 (0.093)	0.395 (0.096)
Num. Post-Treat. Surveys Taken		0.576 (0.029)	0.576 (0.030)
Fin. Index at t1		-0.157 (0.033)	-0.130 (0.037)
Psych. Index at t1		-0.039 (0.037)	-0.071 (0.039)
Cog. Cap. Index at t1		-0.033 (0.033)	-0.005 (0.035)
Health Index at t1		-0.058 (0.037)	-0.056 (0.038)
Online Platform		-0.172 (0.061)	-0.196 (0.063)
Female			-0.163 (0.099)
Age			-0.021 (0.004)
Non-White			-0.139 (0.079)
More than HS			0.102 (0.068)
HH Size			-0.020 (0.021)
Parent			0.567 (0.096)
Partner			0.038 (0.068)
Unearned Inc. at t1 (\$000's)			0.044 (0.047)
Under FPL in 2019			0.020 (0.073)
Wave	No	Yes	Yes
Constant	-0.390 (0.036)	-2.043 (0.239)	-0.746 (0.248)
<i>N</i>	5243	5234	4957
Pseudo <i>R</i> ²	0.010	0.130	0.138

Table H.2: Logistic regressions. Dependent variable: binary indicator for giving access to at least one bank account. “Num. Post-Treat. Surveys Taken” excludes the first Wave 6 t2 survey; thus, for all participants the maximum number of surveys they could have taken is three. “Online Platform” is a binary indicator for having access to the additional features of the online platform. FPL=Federal poverty line. *SEs* are robust and in parentheses.

Effect of UCTs on Bank Balances (\$).

	(1) Bank Bal. Days 0-13	(2) Bank Bal. Days 14-27	(3) Bank Bal. Days 28-41	(4) Bank Bal. Days 42-55
\$500 Group	42.98 (19.79)	7.44 (21.46)	2.10 (29.64)	-11.86 (36.04)
\$2,000 Group	212.76 (38.69)	125.70 (40.82)	49.32 (49.64)	23.73 (54.44)
Pre-UCT Balance (\$)	1.18 (0.04)	0.98 (0.05)	0.98 (0.07)	1.08 (0.08)
Wave	Yes	Yes	Yes	Yes
Constant	245.43 (331.70)	122.14 (184.90)	393.02 (322.33)	185.38 (162.33)
<i>N</i>	1220	1109	1080	1034
<i>R</i> ²	0.696	0.555	0.435	0.430

Table H.3: OLS regressions. DVs: average of bank balances for every two week period after the UCT payment date, from day 0 to day 55. Pre-UCT balance refers to the participant's average daily balance before the UCT date. *SEs* (in parentheses) are robust.

Effect of UCTs on Spending (\$).

	(1) Spending Days 0-13	(2) Spending Days 14-27	(3) Spending Days 28-41	(4) Spending Days 42-55
\$500 Group	18.15 (3.98)	3.02 (4.01)	-0.83 (4.34)	2.50 (6.04)
\$2,000 Group	77.93 (6.16)	9.36 (4.90)	8.35 (5.61)	2.71 (7.77)
Pre-UCT Spending (\$/Day)	0.70 (0.03)	0.66 (0.03)	0.61 (0.03)	0.73 (0.05)
Wave	Yes	Yes	Yes	Yes
Constant	21.68 (41.49)	20.60 (21.20)	30.58 (20.80)	38.07 (26.99)
<i>N</i>	1589	1475	1369	1352
<i>R</i> ²	0.373	0.301	0.291	0.273

Table H.4: OLS regressions. Dependent variables: average of daily spending for every two week period after the UCT payment date, from day 0 to day 55. *SEs* (in parentheses) are robust.

Robustness & Multiple Hypothesis Testing Corrections, Collapsing Across Time Periods.

Model	Outcome	Treatment	Beta	Unadjusted <i>p</i> -value	BH Adjusted <i>p</i> -value	WY Adjusted <i>p</i> -value
Main	Fin.	\$500	-.096	<.0001	<.0001	<.0001
Main	Fin.	\$2,000	-.058	.0473	.0631	.1388
Main	Psych.	\$500	-.109	<.0001	<.0001	<.0001
Main	Psych.	\$2,000	-.130	<.0001	<.0001	<.0001
Main	Cog. Cap.	\$500	-.049	.0924	.0924	.1388
Main	Cog. Cap.	\$2,000	-.070	.0613	.0701	.1388
Main	Health	\$500	-.122	<.0001	<.0001	<.0001
Main	Health	\$2,000	-.143	<.0001	<.0001	<.0001
TOT	Fin.	\$500	-.041	.1506	.2410	.4816
TOT	Fin.	\$2,000	-.034	.3886	.4829	.7696
TOT	Psych.	\$500	-.096	.0013	.0056	.0107
TOT	Psych.	\$2,000	-.104	.0120	.0319	.0684
TOT	Cog. Cap.	\$500	.029	.4226	.4829	.7696
TOT	Cog. Cap.	\$2,000	-.019	.7141	.7141	.7696
TOT	Health	\$500	-.095	.0014	.0056	.0107
TOT	Health	\$2,000	-.078	.0351	.0702	.1634
TOT + AAS	Fin.	\$500	-.027	.4176	.5409	.8132
TOT + AAS	Fin.	\$2,000	-.041	.3528	.5409	.8132
TOT + AAS	Psych.	\$500	-.095	.0050	.0401	.0375
TOT + AAS	Psych.	\$2,000	-.097	.0421	.1122	.2164
TOT + AAS	Cog. Cap.	\$500	.012	.7782	.7782	.8132
TOT + AAS	Cog. Cap.	\$2,000	-.042	.4733	.5409	.8132
TOT + AAS	Health	\$500	-.087	.0110	.0441	.0736
TOT + AAS	Health	\$2,000	-.045	.2916	.5409	.8105
Add'l Covars.	Fin.	\$500	-.096	<.0001	<.0001	<.0001
Add'l Covars.	Fin.	\$2,000	-.066	.0255	.0291	.0516
Add'l Covars.	Psych.	\$500	-.121	<.0001	<.0001	<.0001
Add'l Covars.	Psych.	\$2,000	-.135	<.0001	<.0001	<.0001
Add'l Covars.	Cog. Cap.	\$500	-.067	.0171	.0228	.0507
Add'l Covars.	Cog. Cap.	\$2,000	-.074	.0398	.0398	.0516
Add'l Covars.	Health	\$500	-.123	<.0001	<.0001	<.0001
Add'l Covars.	Health	\$2,000	-.131	<.0001	<.0001	<.0001
No Wave 6	Fin.	\$500	-.117	<.0001	<.0001	<.0001
No Wave 6	Fin.	\$2,000	-.097	.0078	.0104	.0234
No Wave 6	Psych.	\$500	-.133	<.0001	<.0001	<.0001
No Wave 6	Psych.	\$2,000	-.171	<.0001	<.0001	<.0001
No Wave 6	Cog. Cap.	\$500	-.045	.1865	.1865	.1886
No Wave 6	Cog. Cap.	\$2,000	-.084	.0703	.0804	.1388
No Wave 6	Health	\$500	-.145	<.0001	<.0001	<.0001
No Wave 6	Health	\$2,000	-.172	<.0001	<.0001	<.0001
Z Scores	Fin.	\$500	-.081	<.0001	<.0001	<.0001
Z Scores	Fin.	\$2,000	-.048	.0118	.0157	.0207
Z Scores	Psych.	\$500	-.096	<.0001	<.0001	<.0001
Z Scores	Psych.	\$2,000	-.111	<.0001	.0001	<.0001
Z Scores	Cog. Cap.	\$500	-.037	.0372	.0372	.0646
Z Scores	Cog. Cap.	\$2,000	-.049	.0334	.0372	.0646
Z Scores	Health	\$500	-.080	<.0001	<.0001	<.0001
Z Scores	Health	\$2,000	-.096	<.0001	<.0001	<.0001

Table H.5: Survey index outcomes when collapsing across all endline measures: Beta coefficients, unadjusted *p*-values, Benjamini-Hochberg (BH) adjusted *p*-values, and Westfall-Young (WY) adjusted *p*-values. Statistically significant *p*-values are in bold, with unadjusted and WY adjusted *p*-values using a standard $\alpha = 0.05$. Values are from OLS regressions, regressing the index on dummies for the treatment groups while controlling for the baseline value of that index and access to the online platform. All regressions use robust *SEs* clustered at the participant level. Calculations were conducted by model (i.e., eight outcomes at a time). BH adjustments use a *Q* value of 10%. WY corrections were performed using 10,000 bootstraps clustered at the participant level. We use the software provided by Jones, Molitor and Reif (2019). TOT=treatment on the treated, AAS=answered all surveys, Add'l Covars=includes additional covariates (see text).

Multiple Hypothesis Testing Corrections, Splitting by Time Period.

Model	Outcome	Treatment	t	Beta	Unadjusted p-value	BH Adjusted p-value	WY Adjusted p-value
Disaggregated t	Fin.	\$500	2	-.110	.0001	.0004	.0017
Disaggregated t	Fin.	\$500	3	-.074	.0076	.0140	.0904
Disaggregated t	Fin.	\$500	4	-.106	.0008	.0018	.0125
Disaggregated t	Fin.	\$2,000	2	-.077	.0339	.0507	.2578
Disaggregated t	Fin.	\$2,000	3	-.040	.2652	.2893	.5924
Disaggregated t	Fin.	\$2,000	4	-.059	.1506	.1721	.4563
Disaggregated t	Psych.	\$500	2	-.115	.0001	.0003	.0009
Disaggregated t	Psych.	\$500	3	-.087	.0028	.0055	.0375
Disaggregated t	Psych.	\$500	4	-.127	.0002	.0006	.0035
Disaggregated t	Psych.	\$2,000	2	-.079	.0371	.0507	.2589
Disaggregated t	Psych.	\$2,000	3	-.148	.0002	.0006	.0037
Disaggregated t	Psych.	\$2,000	4	-.163	.0001	.0003	.0007
Disaggregated t	Cog. Cap.	\$500	2	-.002	.9559	.9975	.9985
Disaggregated t	Cog. Cap.	\$500	3	-.061	.0969	.1163	.3841
Disaggregated t	Cog. Cap.	\$500	4	-.087	.0287	.0459	.2545
Disaggregated t	Cog. Cap.	\$2,000	2	.000	.9976	.9976	.9985
Disaggregated t	Cog. Cap.	\$2,000	3	-.106	.0282	.0459	.2545
Disaggregated t	Cog. Cap.	\$2,000	4	-.104	.0380	.0507	.2589
Disaggregated t	Health	\$500	2	-.099	.0008	.0018	.0125
Disaggregated t	Health	\$500	3	-.107	.0004	.0009	.0060
Disaggregated t	Health	\$500	4	-.166	<.0001	<.0001	<.0001
Disaggregated t	Health	\$2,000	2	-.065	.0742	.0937	.3581
Disaggregated t	Health	\$2,000	3	-.154	.0001	.0003	.0007
Disaggregated t	Health	\$2,000	4	-.214	<.0001	<.0001	<.0001

Table H.6: Survey index outcomes when disaggregating by post-treatment time point: Beta coefficients, unadjusted p -values, Benjamini-Hochberg (BH) adjusted p -values, and Westfall-Young (WY) adjusted p -values. Statistically significant p -values are in bold, with unadjusted and WY adjusted p -values using a standard $\alpha = 0.05$. Values are from OLS regressions that restrict to the given time period, regressing the index on dummies for the treatment groups while controlling for the baseline value of that index and access to the online platform. All regressions use robust SEs . Calculations for correcting p -values were conducted across all 24 outcomes. BH adjustments use a Q value of 10%. WY corrections were performed using 10,000 bootstraps. We use the software provided by Jones, Molitor and Reif (2019).

Predictors of Survey Responsiveness: Indices, Treatment Groups, and Demographics.

	(1) Any	(2) Any	(3) Any	(4) Num.	(5) Num.	(6) Num.
Fin. Index at t1	-0.0386 (0.0387)	-0.0427 (0.0395)	-0.0595 (0.0438)	-0.0089 (0.0168)	-0.0109 (0.0166)	-0.0199 (0.0182)
Psych. Index at t1	-0.0113 (0.0457)	-0.0136 (0.0464)	-0.0241 (0.0486)	-0.0100 (0.0188)	-0.0095 (0.0186)	-0.0068 (0.0191)
Cog. Cap. Index at t1	-0.0277 (0.0403)	-0.0310 (0.0406)	-0.0484 (0.0421)	-0.0059 (0.0163)	-0.0084 (0.0161)	-0.0178 (0.0168)
Health Index at t1	-0.0404 (0.0455)	-0.0508 (0.0455)	-0.0333 (0.0475)	-0.0155 (0.0186)	-0.0223 (0.0181)	-0.0153 (0.0189)
Online Platform	0.0667 (0.0751)	0.1147 (0.0760)	0.1236 (0.0782)	0.0066 (0.0308)	0.0372 (0.0303)	0.0478 (0.0310)
\$500 Group		0.8834 (0.1015)	0.8873 (0.1048)		0.4649 (0.0337)	0.4764 (0.0346)
\$2,000 Group		0.6270 (0.1243)	0.6019 (0.1274)		0.4381 (0.0441)	0.4289 (0.0453)
Female			0.1381 (0.1150)			0.0919 (0.0491)
Age			0.0124 (0.0048)			0.0052 (0.0019)
Non-White			-0.1874 (0.1028)			-0.1395 (0.0386)
More than HS			0.0642 (0.0855)			0.0553 (0.0335)
HH Size			-0.0422 (0.0251)			-0.0261 (0.0103)
Parent			0.1167 (0.1137)			0.0780 (0.0476)
Partner			0.1008 (0.0827)			0.0465 (0.0329)
Unearned Income at t1 (\$000's)			0.0054 (0.0579)			0.0075 (0.0238)
Under FPL 2019			0.0779 (0.0917)			0.0298 (0.0360)
Wave	Yes	Yes	Yes	Yes	Yes	Yes
Constant	2.4695 (0.3953)	2.2150 (0.3952)	1.4008 (0.3238)	1.9820 (0.0983)	1.7985 (0.0979)	1.7362 (0.1141)
<i>N</i>	5234	5234	4957	5234	5234	4957
(Pseudo) <i>R</i> ²	0.020	0.041	0.044	0.021	0.060	0.070

Table H.7: Models 1 to 3: dependent variable is a binary indicator for taking at least one post-treatment survey; logistic regressions. Models 4 to 6: dependent variable is the number of post-treatment surveys taken, excluding the Wave 6 t2 survey that was administered before the UCT (maximum number: 3); OLS regressions. HS=High school, HH=Household, FPL=Federal poverty line. Robust *SEs* in parentheses.

Predictors of Survey Responsiveness: Treatment Group & Baseline Index Interactions.

	(1) Any	(2) Any	(3) Num.	(4) Num.
AnyCash=1	0.8003 (0.0859)	0.7938 (0.0886)	0.4553 (0.0301)	0.4600 (0.0309)
Fin. Index at t1	0.0236 (0.0469)	0.0046 (0.0508)	0.0290 (0.0222)	0.0160 (0.0235)
AnyCash=1 × Fin. Index at t1	-0.2116 (0.0833)	-0.2042 (0.0864)	-0.0947 (0.0325)	-0.0858 (0.0333)
Psych. Index at t1	-0.0419 (0.0545)	-0.0584 (0.0569)	-0.0473 (0.0251)	-0.0393 (0.0256)
AnyCash=1 × Psych Index at t1	0.0987 (0.1031)	0.1215 (0.107)	0.0928 (0.0367)	0.0802 (0.0377)
Cog. Cap. Index at t1	-0.0231 (0.0468)	-0.0454 (0.0486)	-0.0013 (0.0213)	-0.0179 (0.0221)
AnyCash=1 × Cog. Cap. Index at t1	-0.0316 (0.0946)	-0.0122 (0.0963)	-0.0173 (0.0323)	0.0010 (0.0335)
Health Index at t1	-0.0985 (0.0545)	-0.0755 (0.0567)	-0.0303 (0.0248)	-0.0210 (0.0256)
AnyCash=1 × Health Index at t1	0.1682 (0.0982)	0.1471 (0.1026)	0.0195 (0.0355)	0.0130 (0.0369)
Online Platform	0.1184 (0.0761)	0.1268 (0.0783)	0.0382 (0.0303)	0.0486 (0.0310)
Additional Covariates	No	Yes	No	Yes
Constant	2.2204 (0.3954)	1.4046 (0.3249)	1.8020 (0.0971)	1.7340 (0.1142)
<i>N</i>	5234	4957	5234	4957
<i>R</i> ²	0.042	0.045	0.063	0.072

Table H.8: Models 1 and 2: dependent variable is a binary indicator for taking at least one post-treatment survey; logistic regressions. Models 3 and 4: dependent variable is the number of post-treatment surveys taken, excluding the Wave 6 t2 survey that was administered before the UCT (maximum number: 3); OLS regressions. AnyCash is a binary indicator equaling 0 if the participant is in the Control group and 1 if they are in the \$500 or \$2,000 group. All models include wave number. Additional covariates are gender, age, binary indicator for not only identifying as White, binary indicator for having completed more than high school, household size, binary indicator for being a parent, binary indicator for having a partner or spouse, unearned income at t1, and a binary indicator for being under the federal poverty line in 2019. “Online Platform” is a binary indicator for having access to the additional features of the online platform. Robust *SEs* in parentheses.

Lee Bounds.

Index	Treatment	Lower Coeff.	Lower SE	Upper Coeff.	Upper SE	95% CI
Fin.	\$500	-0.324	0.022	0.090	0.023	[-0.360, 0.128]
Fin.	\$2,000	-0.277	0.028	0.046	0.029	[-0.323, 0.094]
Psych.	\$500	-0.334	0.024	0.211	0.023	[-0.373, 0.250]
Psych.	\$2,000	-0.361	0.030	0.105	0.030	[-0.411, 0.155]
Cog. Cap.	\$500	-0.314	0.023	0.253	0.024	[-0.352, 0.292]
Cog. Cap.	\$2,000	-0.324	0.031	0.181	0.031	[-0.376, 0.233]
Health	\$500	-0.340	0.023	0.211	0.024	[-0.378, 0.250]
Health	\$2,000	-0.373	0.030	0.097	0.031	[-0.422, 0.148]

Table H.9: Lee bounds (Lee, 2009) for each of the four index outcomes, collapsing across post-treatment time periods. Coeff.=Coefficient, SE=Standard Error, CI=Confidence Interval.

Effect of UCTs on Survey Indices with Honaker-King Multiple Imputations.

	(1) Fin.	(2) Fin.	(3) Psych.	(4) Psych.	(5) Cog. Cap.	(6) Cog. Cap.	(7) Health	(8) Health
\$500 Group	-0.082 (0.022)	-0.082 (0.021)	-0.069 (0.025)	-0.078 (0.024)	-0.041 (0.025)	-0.048 (0.024)	-0.078 (0.023)	-0.085 (0.023)
\$2,000 Group	-0.060 (0.028)	-0.067 (0.027)	-0.096 (0.030)	-0.103 (0.030)	-0.060 (0.032)	-0.066 (0.031)	-0.107 (0.036)	-0.112 (0.036)
Fin. Index at t1	0.466 (0.012)	0.382 (0.012)						
Psych. Index at t1			0.398 (0.011)	0.382 (0.011)				
Cog. Cap. Index at t1					0.185 (0.010)	0.162 (0.010)		
Health Index at t1							0.389 (0.011)	0.371 (0.011)
Online Platform	-0.011 (0.019)	-0.007 (0.019)	-0.049 (0.024)	-0.047 (0.024)	-0.027 (0.023)	-0.026 (0.023)	-0.030 (0.023)	-0.028 (0.023)
Additional Covariates	No	Yes	No	Yes	No	Yes	No	Yes
Constant	0.019 (0.015)	0.039 (0.098)	0.045 (0.021)	-0.238 (0.105)	0.028 (0.019)	0.296 (0.105)	0.034 (0.017)	-0.099 (0.103)
Observations	15,729	15,729	15,723	15,723	15,702	15,702	15,720	15,720

Table H.10: Primary specification with Honaker and King (2010) multiple imputations for missing values without lag or lead variables. OLS regressions with the four survey outcome indices as dependent variables, collapsing across all post-treatment time periods, and using robust *SEs* clustered at the participant level (shown in parentheses). “Online Platform” is a binary indicator for having access to the additional features of the online platform. Additional covariates are gender, age, binary indicator for not only identifying as White, binary indicator for having completed more than high school, binary indicator for being a parent, binary indicator for having a partner or spouse, and binary indicator for being under the federal poverty line in 2019.

Effect of UCTs on Survey Indices with Mediation by Saliency of Needs.

	(1) Fin.	(2) Fin.	(3) Psych.	(4) Psych.	(5) Cog. Cap.	(6) Cog. Cap.	(7) Health	(8) Health
\$500 Group	-0.089 (0.024)	-0.061 (0.023)	-0.111 (0.025)	-0.035 (0.021)	-0.058 (0.031)	-0.018 (0.029)	-0.126 (0.025)	-0.073 (0.023)
\$2,000 Group	-0.043 (0.032)	-0.021 (0.030)	-0.132 (0.032)	-0.061 (0.027)	-0.086 (0.039)	-0.039 (0.036)	-0.139 (0.031)	-0.091 (0.029)
Fin. Index at t1	0.651 (0.013)	0.599 (0.013)						
Psych. Index at t1			0.624 (0.011)	0.488 (0.011)				
Cog. Cap. Index at t1					0.305 (0.014)	0.235 (0.014)		
Health Index at t1							0.611 (0.012)	0.526 (0.012)
Money Mind - Cost	0.016 (0.008)		-0.051 (0.007)				-0.033 (0.008)	
Overwhelmed Others' Needs	-0.021 (0.007)		-0.106 (0.007)		-0.111 (0.009)		-0.067 (0.007)	
Spending Stress	-0.136 (0.009)		-0.217 (0.009)		-0.135 (0.010)		-0.157 (0.009)	
Online Platform	-0.011 (0.022)	-0.010 (0.021)	-0.017 (0.022)	-0.020 (0.019)	-0.028 (0.027)	-0.030 (0.025)	-0.015 (0.022)	-0.016 (0.020)
Constant	0.018 (0.019)	0.539 (0.046)	0.039 (0.019)	1.363 (0.043)	0.033 (0.023)	0.885 (0.048)	0.028 (0.019)	0.946 (0.047)
Observations	8373	8373	8370	8370	8401	8401	8370	8370
R ²	0.435	0.463	0.400	0.516	0.092	0.152	0.385	0.444

Table H.11: OLS regressions. For t2 to t4 periods; regressions omit the “needs over the next 30 days” and “hypothetical stimulus check spending” variables because they were measured only in t2. Regressions are restricted to participants who have data for all the relevant mediators to ensure comparability within each pair of columns. Models 5 and 6 omit the “money on the mind” variables because of collinearity with the dependent variable. “Online Platform” is a binary indicator for having access to the additional features of the online platform. *SEs* (in parentheses) are robust and clustered at the participant level.

Participants' Beliefs About Factor(s) Determining RCT Benefit Receipt.

Reason	No Cash	\$500 Group	\$2,000 Group	F	p-value
P Characteristics/Circumstance	0.56	0.57	0.57	0.08	0.9185
Researcher Interest	0.08	0.08	0.07	0.20	0.8174
P Survey Answers	0.06	0.08	0.07	1.77	0.1713
Randomness	0.04	0.07	0.07	3.55	0.0290
"People Are Different"	0.04	0.03	0.03	0.81	0.4436
External Circumstance	0.02	0.02	0.03	0.55	0.5765
Do Not Know	0.18	0.18	0.18	0.00	0.9952
Other	0.12	0.11	0.10	1.07	0.3425

Table H.12: $N = 2,374$. Coded responses of participants' answers to the question, "In 1-2 sentences, can you describe why you think some people got one thing in this study and others got something else?" Means for each treatment group. The F test statistic and corresponding p -value refer to a one-way ANOVA testing for the differences across treatment groups. "P Characteristics/Circumstances"=response mentions the participant's characteristics (e.g., race, geographic area) or circumstance (e.g., financial need) as a factor. "Researcher Interest"=mentions the researchers' or non-profit's study interests as a factor. "P Survey Answers"=mentions the participant's survey answers or how they answered the survey (e.g., timeliness, speed) as a factor. "Randomness"=mentions randomness, lottery, luck of the draw, etc. "People Are Different"=makes a general statement about different people being different or having different circumstances or personalities. "External Circumstances"=mentions factors such as funding constraints, first-come-first-serve. "Do Not Know"=participant indicates not knowing. "Other"=anything not captured by the above. Each response could be coded as multiple categories.

I Model

I.1 Introduction

We propose a discrete time economic model to analyze the financial management strategy of an agent who can choose to take a passive (P) or active (A) approach towards the repayment of a stock of debt, denoted by D . The model operates within a three-period setting, where the agent earns income 0, M_1 , and M_2 at times 0, 1, and 2, respectively. It is assumed that $M_2 > M_1$ and all variables are real and non-negative.

At the beginning, the agent intends to make an initial payment towards the debt, denoted by \bar{d} . Subsequently, Nature introduces a possible monetary bonus B , which can take a value of either 0 or $b > 0$, with probabilities $1 - p$ and p respectively. The agent observes the realization of B . Additionally, Nature introduces a negative shock S (which gets added to the debt), which can take a value of either 0 or $s > 0$, with probabilities $1 - q$ and q respectively. All these aspects of the model, including p and q , are known to the agent.

After observing B , and without knowing the value of S , the agent then chooses to either continue with the passive strategy, maintaining the initial debt payment \bar{d} , or switch to an active strategy. However, the agent only learns the actual realization of S if they choose to be active; otherwise, they believe the shock to be \tilde{s} (with probability q). If the agent chooses the active strategy, there is an associated cost a , and they subsequently have the opportunity to re-optimize the payment towards the debt, now denoted by d^* .

There are two types of agents, the benchmark agent who correctly believes $\tilde{s} = s$ and a behavioral type, who optimistically believes that S takes some value $\tilde{s} \in (0, s)$ (conditional on Nature choosing a positive shock, with probability q) instead. The agent derives utility from consumption, and the utility function $u(\cdot)$ is assumed to be concave. Later in the model, $u(\cdot)$ is assumed to take the form of an isoelastic utility function with parameter η .

The agent's objective is to maximize their expected utility. Their problem is characterized by the initial choice of \bar{d} at time 0, the decision to take an active or passive stance in debt management at time 1, and the choice of $d^*(B, S)$ (if they decided to be active), also at time 1. Time 2 captures what happens considerably later when finances must be settled up. During this final period, the agent is always active and is forced to pay a and any remaining debt.

For a visual representation of the decision problem, see Appendix Figure I.12. Note that the visual representation deviates from the exact timing in the model description, but captures the same problem.

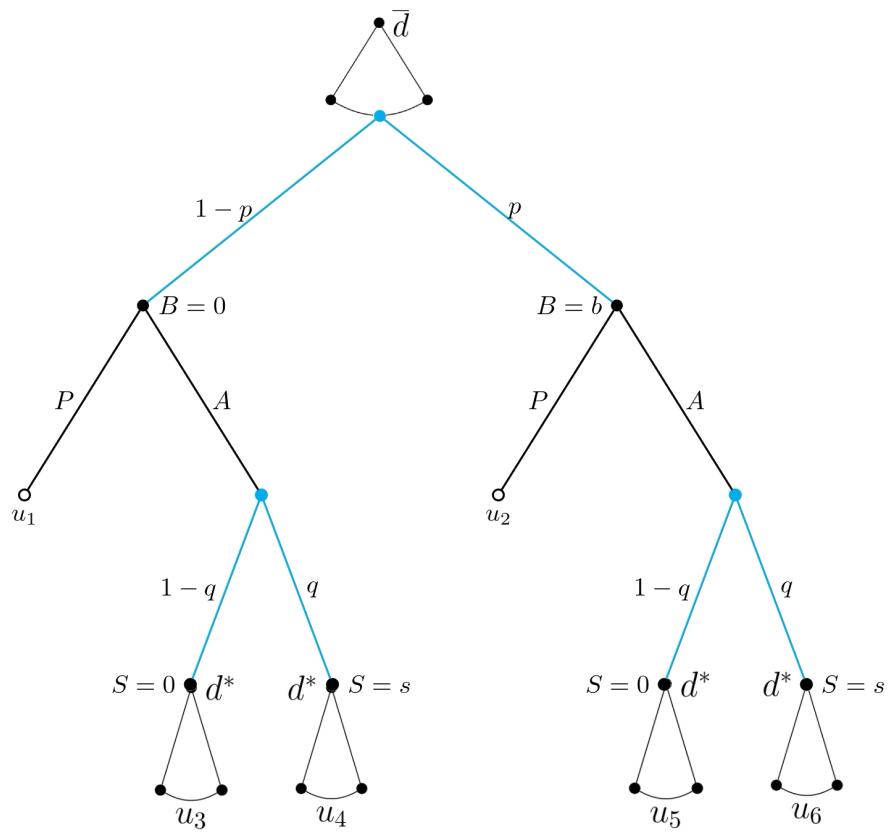


Figure I.12: Extensive-form representation of the model. The agent's decisions are colored black and Nature's are colored light blue.

The explicit payoffs, where d^* takes the appropriate arguments, are as follows:

$$u_1 = u(M_1 - \bar{d}) + \mathbb{E}[u(M_2 - S - a - (D - \bar{d}))] \quad (1)$$

$$u_2 = u(M_1 + b - \bar{d}) + \mathbb{E}[u(M_2 - S - a - (D - \bar{d}))] \quad (2)$$

$$u_3 = u(M_1 - a - d^*) + u(M_2 - a - (D - d^*)) \quad (3)$$

$$u_4 = u(M_1 - a - d^*) + u(M_2 - s - a - (D - d^*)) \quad (4)$$

$$u_5 = u(M_1 + b - a - d^*) + u(M_2 - a - (D - d^*)) \quad (5)$$

$$u_6 = u(M_1 + b - a - d^*) + u(M_2 - s - a - (D - d^*)) \quad (6)$$

I.2 Optimal d^*

Proceeding by backwards induction, we can find the optimal decisions at every point. For each u_i , when $i \geq 3$,

$$d_i^* = \arg \max_{d_i^*} \{u_i\} \quad (7)$$

Recall the assumption that u is concave. To derive the conditions for the optimal d_i^* , we need to set each of the derivatives of the utility functions equal to zero and solve for d_i^* . The equations are:

$$\begin{aligned} \frac{du_3}{dd_3^*} = 0 &\implies u'(M_1 - a - d_3^*) = u'(M_2 - a - (D - d_3^*)) \\ \frac{du_4}{dd_4^*} = 0 &\implies u'(M_1 - a - d_4^*) = u'(M_2 - s - a - (D - d_4^*)) \\ \frac{du_5}{dd_5^*} = 0 &\implies u'(M_1 + b - a - d_5^*) = u'(M_2 - a - (D - d_5^*)) \\ \frac{du_6}{dd_6^*} = 0 &\implies u'(M_1 + b - a - d_6^*) = u'(M_2 - s - a - (D - d_6^*)) \end{aligned}$$

I.2.1 Isoelastic case

For concreteness we now make the standard assumption that u is the isoelastic function with parameter $\eta > 0$:

$$u(x) \equiv \begin{cases} \frac{x^{1-\eta}-1}{1-\eta}, & \eta \neq 1, \\ \ln(x), & \eta = 1. \end{cases} \quad (8)$$

With that in mind, for each i , we can explicitly obtain d_i^* when $u(\cdot)$ is isoelastic.¹²

¹²Note that the table results hold both when $\eta \neq 1$ and when $u(\cdot) = \ln(\cdot)$ for $\eta = 1$.

Isoelastic Utility for any $\eta > 0$	
d_3^*	$\frac{1}{2}(M_1 - M_2 + D)$
d_4^*	$\frac{1}{2}(M_1 - M_2 + s + D)$
d_5^*	$\frac{1}{2}(M_1 - M_2 + b + D)$
d_6^*	$\frac{1}{2}(M_1 - M_2 + b + s + D)$

I.3 Strategy

Conditional on $B = 0$, the agent will weakly prefer to be passive iff

$$\mathbb{E}[P|B = 0] = u_1 \geq \mathbb{E}[A|B = 0] = (1 - q)u_3 + qu_4. \quad (9)$$

Similarly, when $B = b$, the agent will weakly prefer to be passive iff

$$\mathbb{E}[P|B = b] = u_2 \geq \mathbb{E}[A|B = b] = (1 - q)u_5 + qu_6. \quad (10)$$

One natural strategy, having optimally chosen \bar{d} in advance for the case that $B = 0$ (see below), is to be active precisely when the bonus is received and there is more money available. Of course, if a is sufficiently (unrealistically) large then it never makes sense to pay it, while if a is sufficiently small (and especially if a large shock occurs with an intermediate probability and is therefore difficult to plan around) it will always be optimal to pay it. However, for many sets of parameters, including the ones in Appendix Section I.5 inspired by the empirical setting, the agent will indeed rationally be passive iff they do not receive a bonus. That is, in this setting, Equation 10 is always true and Equation 9 is always false.

I.4 Optimal \bar{d}

The choice of \bar{d} will naturally depend on what is expected later: it does not matter if the agent will always be active, and it is crucial if the agent will always be passive. We focus here on the most relevant case for our context, namely when (as above) the agent is passive exactly when they do not get a bonus. Hence, the agent will initially select \bar{d} solely considering the scenario depicted in the leftmost branch in Appendix Figure I.12. In other words, they set

$$\bar{d} = \arg \max_{\bar{d}} \{u_1\}, \quad (11)$$

which we estimate numerically for any given set of applicable parameters.

I.5 Parametrization and simulations

In this section, we investigate which parameters—and in particular what range of a —leads to the strategy $P \iff B = 0$ described in Appendix Section I.3. To run the simulations, we assume the

Parameter	Default Value	Description
M_1	1	Household income for the median participant at a representative period, normalized to one
M_2	2	Income at some later period when they must settle all debts
D	1.5	Debt stock
p	0.01	1% chance of unconditional cash bonus
b	0.5	Size of the bonus
q	0.8	80% chance of experiencing a shock
s	0.8	Shock size according to a rational agent
\tilde{s}	0.24	Shock size according to a behavioral agent
η	1.4	Isoelastic utility parameter, as estimated in 20 OECD countries (Evans, 2005)

Table I.13: Parameters used in the simulations.

parameters in Appendix Table I.13.¹³ In this setting, monetary values can be interpreted roughly as 1 unit = \$1,000.

Having fixed the parameters, we then plot $\mathbb{E}[P|B = 0] - \mathbb{E}[A|B = 0]$ as a function of a , denoted f_1 . Here, \bar{d} and d^* are chosen optimally in the background. When the function is positive, an agent without a bonus chooses to be passive. We do the same for $f_2 \equiv \mathbb{E}[A|B = b] - \mathbb{E}[P|B = b]$. The values of a for which both functions are positive lead to $P \iff B = 0$.

In Appendix Figure I.13, the range of a for which $f_1(\cdot), f_2(\cdot) > 0$ is highlighted in green. The green-shaded area indicates the range of a for which it is optimal to be passive if and only if there is no bonus. If a is too large, the cost of re-optimizing is too high and agents will not choose to do so. Conversely, if a is too small, agents will always choose to be active and observe the shock, trivializing the model.

Our empirical findings suggested a surprising finding, namely that the utility of those who received the bonus was lower than that of those who did not. In Appendix Figure I.14, we plot this difference in their utility in black. Utility in this context refers to the agent's payoff at period 1, which, if they choose P (i.e., when they do not receive a bonus) is u_1 , which involves an expectation of what will happen in period 2. If they choose A (i.e., when they receive a bonus), it is u_5 or u_6 .

In Appendix Figure I.14, the purple line indicates when utilities are equal with and without the bonus. To the right of the purple line, the range of a is shaded in purple where agents, who choose passive iff they receive no bonus, have a worse utility when they received the bonus. The green shading is the same as in the previous figure, where the red line is the minimum value of a for f_1 to be positive and the blue line is the maximum value of a , beyond which f_2 is never positive.

Connecting the results in Appendix Figure I.14 back to our empirical findings, we note that there exists a nontrivial range of a for which agents indeed show more positive utility in Period 1

¹³The full code is available here: https://github.com/arvomm/uct_model

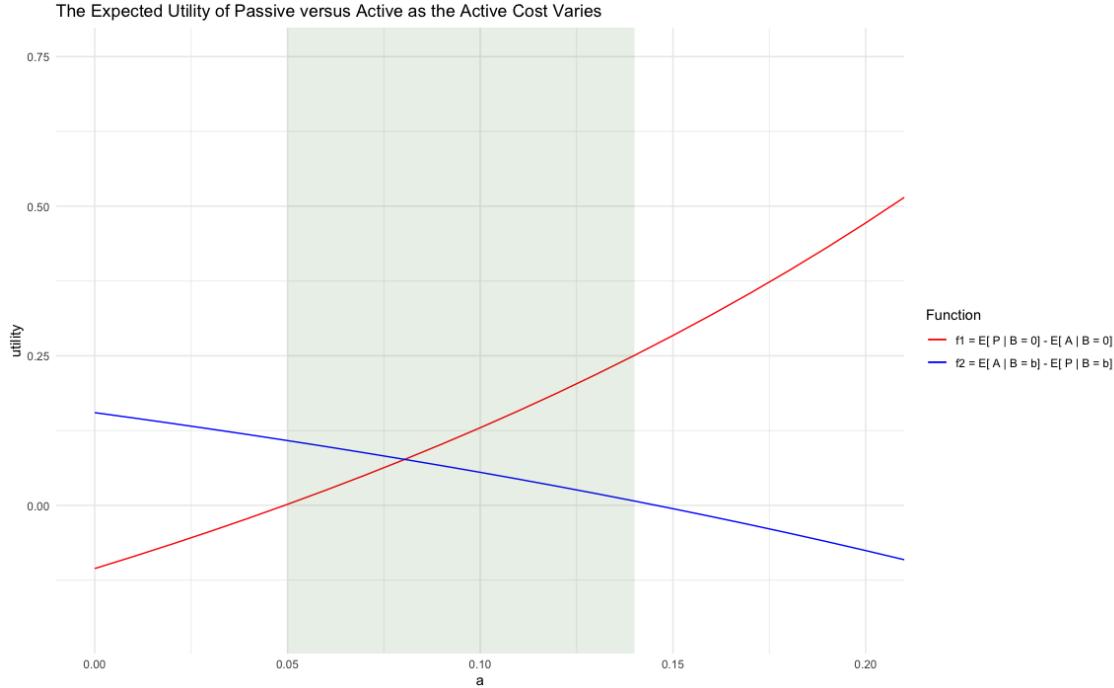


Figure I.13: A plot of f_1 and f_2 varying a on the x-axis, holding constant the other parameters as defined in Appendix Table I.13.

without the bonus than with the bonus. In other words, agents are “worse off” in this period due to the bonus. (However, those agents’ overall “lifetime utility” will be higher due to having received the bonus and choosing the active strategy.) The intuition is that getting some extra money causes people to (rationally) spend effort and resources on dealing with their finances. This may in turn bring unwelcome news about the magnitude of negative shocks that were not previously salient and would otherwise have remained in the background, until they ultimately could not be avoided any longer.

To observe behavior most generally across a range of values of a and \tilde{s} , we plot Appendix Figure I.15 in three dimensions. Here, we can see the utility loss of receiving the bonus for different values of a and \tilde{s} .¹⁴ We only display points for the parameter combinations where it is optimal to be passive iff the agent receives no bonus, as in Appendix Section I.3. (Note that the previous 2-dimensional Appendix Figure I.14 is a slice of Appendix Figure I.15 for a fixed \tilde{s} .)

¹⁴The results are robust to variations in several of the key parameters in the model.

Utility given no bonus minus utility given bonus

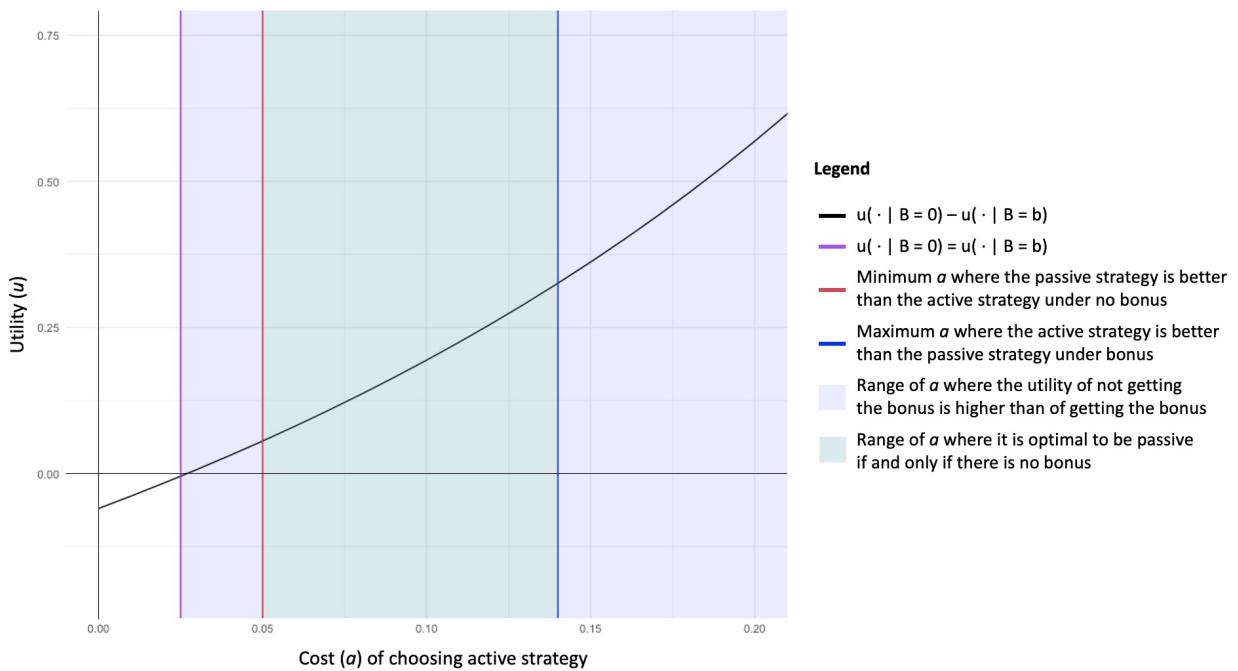


Figure I.14: Utility in period 1 without the bonus, minus utility with the bonus, for the default parameters in Appendix Table I.13.

Utility Difference given Bias and Active Cost

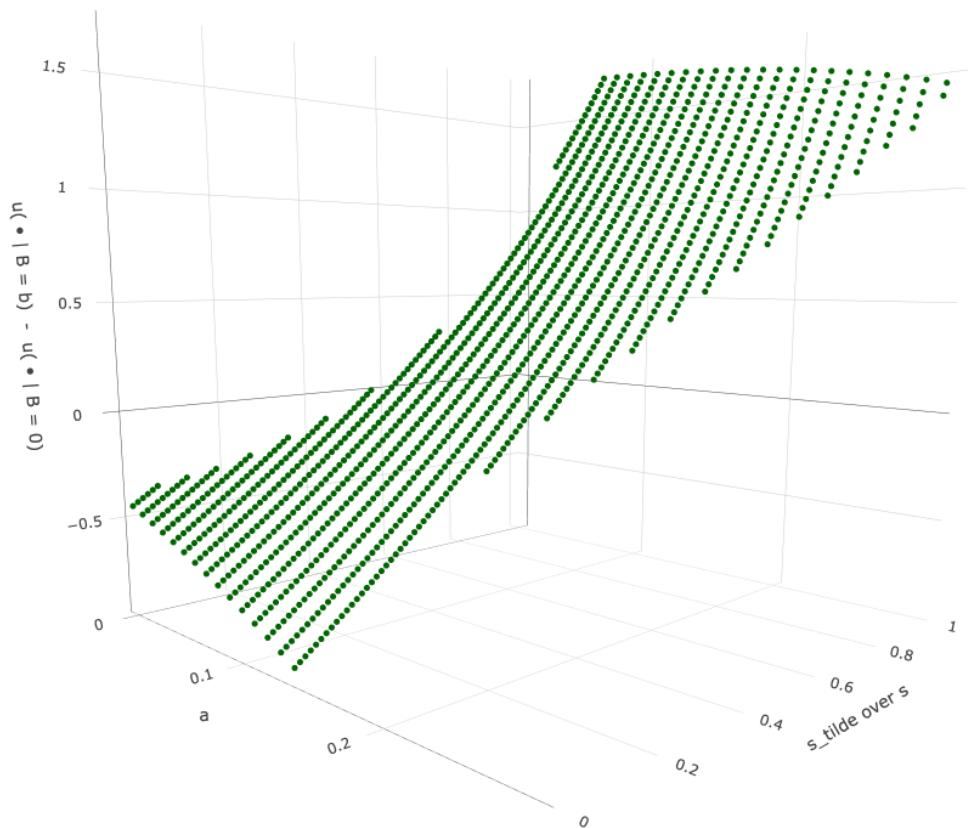


Figure I.15: Utility (period 1) without the bonus, minus utility with the bonus, for various a and \tilde{s} , fixing s and other parameters as in Appendix Table I.13.

References

- Anderson, Michael L.** 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association*, 103(484): 1481–1495.
- Baird, Sarah, and Berk Özler.** 2012. "Examining the Reliability of Self-Reported Data on School Participation." *Journal of Development Economics*, 98(1): 89–93.
- Baird, Sarah, Craig McIntosh, and Berk Özler.** 2019. "When the Money Runs out: Do Cash Transfers Have Sustained Effects on Human Capital Accumulation?" *Journal of Development Economics*, 140: 169–185.
- Banerjee, Abhijit V, and Esther Duflo.** 2007. "The Economic Lives of the Poor." *Journal of Economic Perspectives*, 21(1): 141–167.
- Beegle, Kathleen, Joachim De Weerdt, Jed Friedman, and John Gibson.** 2012. "Methods of Household Consumption Measurement through Surveys: Experimental Results from Tanzania." *Journal of Development Economics*, 98(1): 3–18.
- Brune, Lasse, Xavier Giné, Jessica Goldberg, and Dean Yang.** 2017. "Savings Defaults and Payment Delays for Cash Transfers: Field Experimental Evidence from Malawi." *Journal of Development Economics*, 129: 1–13.
- Cohen, Jacob.** 1988. *Statistical Power Analysis for the Behavioral Sciences*. . Second ed., Lawrence Erlbaum Associates.
- Dana, Jason, Roberto A. Weber, and Jason Xi Kuang.** 2007. "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness." *Economic Theory*, 33(1): 67–80.
- DellaVigna, Stefano, and Devin Pope.** 2018. "Predicting Experimental Results: Who Knows What?" *Journal of Political Economy*, 126(6): 2410–2456.
- DellaVigna, Stefano, Devin Pope, and Eva Vivalt.** 2019. "Predict Science to Improve Science." *Science*, 366(6464): 428–429.
- Evans, David J.** 2005. "The Elasticity of Marginal Utility of Consumption: Estimates for 20 OECD Countries*." *Fiscal Studies*, 26(2): 197–224.
- Evans, David K., and Anna Popova.** 2017. "Cash Transfers and Temptation Goods." *Economic Development and Cultural Change*, 65(2): 189–221.
- Frederick, Shane, George Loewenstein, and Ted O'Donoghue.** 2002. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature*, 40(2): 351–401.
- Gerber, Alan S.** 2012. "Attrition." In *Field Experiments: Design, Analysis, and Interpretation*. . First ed., , ed. Alan S. Gerber and Donald P. Green, 211–252. New York, NY, US:W. W. Norton.
- Ghanem, Dalia, Sarojini Hirshleifer, and Karen Ortiz-Beccera.** 2023. "Testing Attrition Bias in Field Experiments." *Journal of Human Resources*.
- Godoy, Ricardo, Dean Karlan, and Jonathan Zinman.** 2021. "Randomization for Causality, Ethnography for Mechanisms: Illiquid Savings for Liquor in an Autarkic Society." *National Bureau of Economic Research*, w29566.
- Gross, Tal, and Jeremy Tobacman.** 2014. "Dangerous Liquidity and the Demand for Health Care Evidence from the 2008 Stimulus Payments." *Journal of Human Resources*, 49(2): 424–445.
- Haushofer, Johannes, and Ernst Fehr.** 2014. "On the Psychology of Poverty." *Science*, 344(6186): 862–867.
- Honaker, James, and Gary King.** 2010. "What to Do about Missing Values in Time-Series Cross-Section Data." *American Journal of Political Science*, 54(2): 561–581.
- Horowitz, Joel L., and Charles F. Manski.** 2000. "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data." *Journal of the American Statistical Association*

- Association*, 95(449): 77–84.
- Jones, Damon, David Molitor, and Julian Reif.** 2019. “What Do Workplace Wellness Programs Do? Evidence from the Illinois Workplace Wellness Study.” *The Quarterly Journal of Economics*, 134(4): 1747–1791.
- Landis, J. Richard, and Gary G. Koch.** 1977. “The Measurement of Observer Agreement for Categorical Data.” *Biometrics*, 33(1): 159–174.
- Lasky-Fink, Jessica, and Elizabeth Linos.** 2022. “It’s Not Your Fault: Reducing Stigma Increases Take-up of Government Programs.” *SSRN Electronic Journal*.
- Lee, David S.** 2009. “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects.” *The Review of Economic Studies*, 76: 1071–1102.
- Loewenstein, George F., and Dražen Prelec.** 1993. “Preferences for Sequences of Outcomes.” *Psychological Review*, 100: 91–108.
- Martinelli, César, and Susan Wendy Parker.** 2009. “Deception and Misreporting in a Social Program.” *Journal of the European Economic Association*, 7(4): 886–908.
- McNeill, Kristen, and Rachael Pierotti.** 2021. “Reason-Giving for Resistance: Obfuscation, Justification and Earmarking in Resisting Informal Financial Assistance.” *Socio-Economic Review*.
- Moore, Jeffrey C, Linda L Stinson, and Edward J Welniak.** 2000. “Income Measurement Error in Surveys: A Review.” *Journal of Official Statistics*, 16(4): 331–361.
- O’Brien, Rourke L.** 2012. “Depleting Capital? Race, Wealth and Informal Financial Assistance.” *Social Forces*, 91(2): 375–396.
- Portes, Alejandro.** 1998. “Social Capital: Its Origins and Applications in Modern Sociology.” *Annual Review of Sociology*, 25.
- Rubin, Donald B.** 2004. *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ:John Wiley & Sons.
- Shah, Anuj K., Jiaying Zhao, Sendhil Mullainathan, and Eldar Shafir.** 2018. “Money in the Mental Lives of the Poor.” *Social Cognition*, 36(1): 4–19.
- Snowberg, Erik, and Leeat Yariv.** 2021. “Testing the Waters: Behavior across Participant Pools.” *American Economic Review*, 111(2): 687–719.
- Yoo, Paul Y., Greg J. Duncan, Katherine Magnuson, et al.** 2022. “Unconditional Cash Transfers and Maternal Substance Use: Findings from a Randomized Control Trial of Low-Income Mothers with Infants in the U.S.” *BMC Public Health*, 22(1): 897.