NBER WORKING PAPER SERIES

IS MONEY OVERRATED?
MISPERCEIVED SATISFACTION FROM INCOME

Rafael Macedo-Rubião
Ricardo Perez-Truglia

Working Paper 35423
http://www.nber.org/papers/w35423

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2026, Revised November 2026

We are thankful for comments by Daniel Benjamin, Claudia Senik, Andrew Clark, Andrew Oswald, and seminar participants at UCLA-Anderson and Universidad de San Andres. The experiment was pre-registered at the AEA RCT Registry (AEARCTR-0018251) and was reviewed and approved in advance by the Institutional Review Board at UCLA. We are grateful for financial support from the Behavioral Lab at UCLA's Anderson School of Management. We used artificial intelligence (AI) tools to assist with survey development, data analysis, proof writing, and copyediting; we remain responsible for all content and any errors. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2026 by Rafael Macedo-Rubião and Ricardo Perez-Truglia. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Is Money Overrated? Misperceived Satisfaction from Income
Rafael Macedo-Rubião and Ricardo Perez-Truglia
NBER Working Paper No. 35423
July 2026, Revised November 2026
JEL No. C93, D91, I31

ABSTRACT

People often make important choices in pursuit of higher income, but do they place too much weight on income when deciding what path to take? We test the hypothesis that individuals overestimate the marginal satisfaction from income—the expected effect of an income increase on overall life satisfaction—and therefore give income too much importance in their decisions. We designed a pre-registered experiment that elicits beliefs about the marginal satisfaction from income and randomly assigns respondents to receive scientific evidence about it. Respondents report beliefs about the marginal satisfaction from income that are substantially above the scientific benchmark, both for themselves and for others, with a larger gap for themselves. These gaps shrink when respondents are exposed to scientific evidence, and the effects persist one month later. To study whether these beliefs shape decisions, we examine how respondents weigh higher income against other considerations, such as having to work longer hours. We measure these trade-offs using personalized job-choice scenarios and a real-world decision each respondent is facing. To elicit the latter, we make a methodological contribution: an AI-led interview method that moves beyond static, predefined survey instruments by combining the flexibility of qualitative interviewing with the discipline of closed-ended survey measurement. We find that these beliefs are consequential: after learning that income matters less for life satisfaction than they initially thought, respondents place less weight on income in their decisions.

Rafael Macedo-Rubião
University of California, Los Angeles
rafael.rubiao.phd@anderson.ucla.edu

Ricardo Perez-Truglia
University of California, Los Angeles
and NBER
ricardotruglia@gmail.com

An online appendix is available at
http://www.nber.org/data-appendix/w35423
A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/18251

1 Introduction

People routinely make important decisions that require trading off income against other things: whether to work longer hours, take a new job, or choose a high-earning career. These decisions implicitly require answering questions such as: how would I feel if I had more money but had to work longer hours? Or more money but less time with my family? The difficulty is that people observe only the life they choose, not the counterfactual. Someone who chooses the extra hours does not know how they would have felt had they chosen more leisure instead. Still, individuals can draw on experience to form beliefs about the effects of income on their well-being and use those beliefs to guide future decisions. However, biases in these beliefs could lead to suboptimal choices. We investigate this possibility through a pre-registered survey experiment that measures these beliefs, examines how they respond to scientific evidence, and tests whether they influence economic choices.

A key challenge for our empirical approach is how to measure individuals' beliefs about the benefits of higher income. Economists model these decisions as trade-offs between utility gains and losses. For example, an individual deciding whether to work an additional hour weighs the utility gain from extra earnings against the utility loss from additional work. Ideally, we would directly measure people's perceptions of these marginal gains and losses. Utility, however, is an abstract representation of preferences with no natural measurement scale, so we cannot directly elicit perceptions of marginal utility. Instead, we draw on the long tradition in the social sciences of measuring life satisfaction and other forms of subjective well-being. Importantly, our framework does not require life satisfaction to be equivalent to utility. In other words, individuals make choices to maximize utility, and life satisfaction is just one of many inputs into their utility function (Becker and Rayo, 2008; Benjamin et al., 2014).

Our analysis centers on beliefs about the *marginal satisfaction from income* (MSI). We define the *self-MSI* as the effect of a 20% increase in one's own income on one's own life satisfaction, measured on a scale from 0 (extremely dissatisfied) to 100 (extremely satisfied). For example, a self-MSI belief of 5 means that the respondent expects such an income increase to raise their life satisfaction by 5 points. Our central question is whether individuals hold accurate beliefs about their self-MSI. If they overestimate it, they may place excessive weight on income when making trade-offs between higher earnings and other sources of satisfaction, such as family time or leisure. We also elicit beliefs about the *average-MSI*, defined as the effect of a 20% increase in the average person's income on that person's life satisfaction. In the empirical analysis, we use prior and posterior MSI as shorthand for these beliefs before and after the information-treatment stage. Although only self-MSI beliefs should directly

influence the income-related choices we study, scientific evidence about average-MSI provides a way to experimentally shift these beliefs.

To guide the experimental design, we use a simple conceptual framework for understanding how individuals form beliefs about the marginal satisfaction from income. The basic idea is that individuals learn from their own experiences and those of others. For example, someone may recall how happy they felt after buying a new car. That initial excitement, however, may fade as they get used to the car. If individuals fail to fully account for this adaptation, they may overestimate the lasting satisfaction gains. The framework yields three simple predictions that we test in our experiment. First, individuals overestimate both self-MSI and average-MSI, and more so the former. Second, providing scientific evidence about average-MSI lowers beliefs about both average-MSI and self-MSI, with greater updating for the former. Third, by lowering self-MSI beliefs, this information makes individuals less inclined to choose higher-income options.

To test these predictions, we designed a pre-registered online information-provision experiment. The baseline survey begins by eliciting respondents' prior beliefs about average-MSI and self-MSI, after which respondents are randomly assigned to treatment or control. Treated respondents receive scientific evidence about average-MSI, drawn from correlational and causal studies, indicating that a 20% increase in income raises life satisfaction by about one point on a 0–100 scale (i.e., average-MSI ≈ 1). The treatment also briefly discusses why this effect may be modest, including adaptation and rising aspirations. We then re-elicit beliefs about both average-MSI and self-MSI to measure how respondents update each. Finally, to assess whether these changes in beliefs affect choices, we present respondents with decisions involving trade-offs between income and other considerations, such as working longer hours for higher pay.

We first measure these trade-offs using three job-choice scenarios tailored to each respondent. To construct these scenarios, we collected information about each respondent's current or most recent job before the information treatment. In the first scenario, respondents weigh higher earnings against a longer commute; in the second, higher earnings against longer work hours; and in the third, higher earnings against fewer hours of sleep. For each scenario, we ask respondents which option they would choose (the "choice framing"), with responses coded on a 1–4 scale so that higher values indicate a greater preference for the higher-income option. Additionally, following Benjamin et al. (2014), we ask respondents which option they expect would make them more satisfied with their lives (the "satisfaction framing").[1]

Although tailored to each respondent, the job-choice scenarios were hypothetical. To complement them, we also elicited preferences over real-world decisions respondents were actually

[1]The two sets of questions were presented in randomized order to mitigate order effects.

facing. To do so, we developed a mixed-methods procedure that uses an AI-led interview to construct a question tailored to each respondent's decision. We first asked respondents to identify a decision they were considering that involved a trade-off between income and another consideration. Respondents then chatted with an AI interviewer, who gathered the relevant details of their situation. Using this information, the interviewer constructed a question presenting a choice between Option A and Option B, following the same format and 1–4 scale as the job-choice scenarios. For example, one respondent was deciding whether to move to a different city (Option A) or stay in their current city (Option B), and explained that moving would offer a higher-paying job but would also mean living farther away from friends. Another respondent was deciding whether to work extra hours to help pay for an expensive wedding, at the cost of less time with family and a greater risk of burnout.

To measure the persistence of treatment effects and track respondents' real-world decisions, we designed a follow-up survey to be administered about one month after the baseline. The follow-up re-elicits beliefs about self-MSI and average-MSI, repeats the job-choice scenarios, and asks whether respondents have made the real-world decision discussed at baseline and, if so, which option they chose.

In April 2026, 3,002 respondents recruited via Prolific completed the baseline survey. The survey included multiple quality checks. After applying sample restrictions, including the exclusion of respondents with extreme prior beliefs, our main analysis sample at baseline consists of 2,775 respondents. All were invited to participate in the follow-up survey, and about 82.3% completed it. The treatment and control groups are well balanced on pre-treatment characteristics and have similar follow-up response rates. As is common in online samples, respondents are somewhat younger and more educated than the general U.S. population.

We begin our analysis by describing the real-world decisions respondents reported and explaining how we code them. These decisions involve a wide range of trade-offs. On the income side, respondents most often considered taking higher-paying jobs or working more hours, followed by decisions such as relocating, freelancing, or returning to work. On the non-income side, these financial benefits were most often weighed against family time, free time, schedule flexibility, and proximity to family and friends. We code responses to the real-world decisions on the same 1–4 scale used for the hypothetical job choices, with higher values indicating greater willingness to choose the higher-income option. Unlike the hypothetical scenarios, however, the real-world decisions were generated in real time, so neither the nature of the trade-off nor which option offered higher income was known in advance. For example, working extra hours could appear as either Option A or Option B, depending on the respondent. We therefore use a large language model (LLM) to classify these decisions ex

post, identifying which option is income-increasing. Our main analysis focuses on the subset of 1,711 decisions that respondents rated as accurate and for which the higher-income option can be reliably identified.[2]

We next examine respondents' prior beliefs about the marginal satisfaction from income. Mean prior average-MSI is 5.20. That is, on average, respondents believe that a 20% income increase would raise the average person's life satisfaction by 5.20 points on the 0–100 scale. Consistent with the view that respondents have limited information about average-MSI, there is considerable disagreement in their prior beliefs, and most respondents express some uncertainty about these beliefs when asked in the follow-up survey. Mean prior self-MSI is higher, at 6.80, and most respondents also express uncertainty about this belief. We then provide suggestive evidence that these beliefs contain signal, even if they are noisy. First, we show that these beliefs are highly persistent over time. In the control group, where respondents reported their beliefs twice during the baseline survey without receiving any new information, we estimate test-retest correlations of 0.729 for average-MSI and 0.876 for self-MSI. Using the follow-up survey, we also measure correlations between beliefs elicited one month apart, finding correlations of 0.345 and 0.457, respectively. Finally, we show that prior beliefs have power to predict choices: respondents who report lower MSI beliefs tend to place less weight on income when making choices.

Our experimental findings are consistent with the three key predictions of the conceptual framework. First, we find that—relative to the scientific benchmark—the average respondent holds upward-biased beliefs about the marginal satisfaction from income.[3] Mean prior average-MSI is 5.20 points, roughly five times the scientific benchmark of 1 point. Under the additional assumption that this benchmark is an unbiased estimate for our subject pool as well as the general population—see Section 5.2 for a discussion—we find an even larger upward bias in self-MSI beliefs, with a mean prior belief of 6.80 points, almost seven times the benchmark.

Second, when presented with new information, respondents revise both their self-MSI and average-MSI beliefs toward the scientific evidence. For self-MSI, the mean posterior belief is 4.23 in the treatment arm, compared with 6.95 in the control arm ($p<0.001$). For average-MSI, the corresponding means are 2.50 and 5.81, respectively ($p<0.001$). As expected, updating is stronger for average-MSI than for self-MSI, since the treatment provides direct evidence about the former but is only indirectly informative about the latter. Moreover, these belief changes are persistent, with about one-third of the treatment effects remaining

[2]Some excluded decisions, for example, involve trade-offs between immediate and future financial benefits, rather than between income and a non-income consideration.

[3]For brevity, we use "upward-biased" and related language as shorthand for beliefs that exceed the scientific benchmark used in the experiment.

one month later, suggesting that respondents retained some of what they learned.

Third, we find some statistically significant effects of the information treatment on choices in the expected direction. Relative to the control arm, treated subjects (who receive evidence that the average-MSI is lower than they thought) become less willing to choose the higher-income option. For example, consider the scenario involving 20% higher earnings at the expense of 20% more work hours. Respondents in the treatment arm reduced their willingness to choose the higher-income option by about 0.089 SD (p=0.020). The effects are close in size for the other two job-choice scenarios and the real-world decision and are mostly similar across the choice framing and the satisfaction framing. However, due to lower statistical precision, these coefficients are not always statistically significant. Still, the consistent pattern across outcomes and framings supports the interpretation that reducing individuals' self-MSI beliefs shifts them away from higher-income options.

To put the magnitude of the effects on behavior in perspective, we also look at the correlation between self-MSI beliefs and decision-making. As a first step, we estimate a simple linear regression of willingness to choose the higher-income option on the respondent's posterior self-MSI. This analysis shows a positive, statistically significant relationship between posterior self-MSI and pro-income choices. In other words, when facing a choice with a trade-off between income and a non-income consideration (e.g., leisure), respondents with a higher posterior self-MSI are more willing to choose the higher-income option. This is true for each of the three job-choice scenarios as well as for the real-world decisions. These associations are also economically meaningful. For example, for the choice trading off higher income versus fewer work hours, a 1-point reduction in the respondent's posterior self-MSI is significantly (p<0.001) associated with a reduction of 0.023 standard deviations (SD) in willingness to choose the higher-income option. The estimates are similar in size for each of the other two job-choice scenarios and for the real-world decision, ranging from 0.020 to 0.030 SD and remaining statistically significant. Estimates under the satisfaction framing are also similar.

The OLS specification uses both experimental and non-experimental variation in beliefs. While this approach maximizes statistical power by using all available variation, the estimates could suffer from omitted-variable bias. To estimate the causal effect of beliefs on choices, we use a simple two-stage least squares (2SLS) model that estimates the effect of beliefs about one's own marginal satisfaction from income on choices by using only the variation in beliefs driven by random assignment to the information treatment. The results from the 2SLS model are similar in magnitude to the results from the OLS model discussed above, but less precisely estimated—which is expected, since it can use only part of the overall variation in beliefs. For the choice trading off higher income for fewer work hours, a 1-point reduction in respondents' posterior self-MSI is associated with a 0.033 SD (p=0.020) reduction in their

willingness to choose the higher-income option. The 2SLS point estimate of 0.033 SD is somewhat larger in magnitude than the corresponding OLS estimate of 0.023 SD, but we cannot reject the null hypothesis that they are equal. While the 2SLS estimates for the other choices and the alternative framing are not always statistically significant, they all show the same sign and expected order of magnitude, which is reassuring.

Our paper relates to and contributes to several strands of literature. First, we relate to the extensive literature on the effects of income on subjective well-being (Clark et al., 2018). These studies focus on estimating the actual causal effect of income on life satisfaction (Apouey and Clark, 2015; Haushofer and Shapiro, 2016; Lindqvist et al., 2020; Dwyer and Dunn, 2022) and the underlying mechanisms, such as hedonic adaptation (Di Tella et al., 2010; Clark and Georgellis, 2013), income aspirations (Stutzer, 2004), and income comparisons (Luttmer, 2005; Clark et al., 2008). The contribution of this paper is to shift the focus from the *actual* marginal satisfaction from income, which has been the focus of the existing literature, to the *perceived* marginal satisfaction from income. We are the first to measure these perceptions and to estimate their causal effects on economic behavior.[4]

This paper also relates to the longstanding literature in psychology and economics on imperfect self-knowledge and affective forecasting errors. Seminal contributions include Wilson and Gilbert (2003) on affective forecasting, Loewenstein et al. (2003) on projection bias, and Kahneman and Thaler (2006) on experienced versus decision utility. Existing empirical studies in this literature often ask respondents to predict their future happiness after a life event and later compare those predictions with realized happiness. This type of evidence has two limitations that may help explain why it has received less attention in economics. First, it is not causal. We address this limitation by developing an experimental framework that identifies causal effects. Second, from an economic perspective, forecasting errors are relevant only insofar as they lead to suboptimal choices, yet the existing evidence does not measure behavior. We contribute to this literature by developing an empirical framework that links misperceptions directly to economic decision-making.

Our paper also contributes to an emerging literature that uses LLMs for survey measurement. Our methodological contribution is to use AI-led interviews to combine qualitative and quantitative methods. Recent studies propose using LLMs to conduct interviews and collect qualitative data at scale (Chopra and Haaland, 2023; Geiecke and Jaravel, 2026).[5] We use AI-led interviews for a different purpose: the AI interviewer first interviews respondents

[4]For the first part—measuring the perceived effect of income on happiness—there are two exceptions, both from the psychology literature: Aknin et al. (2009) asked respondents to predict the happiness of others, and their own happiness, at different income levels; and Cone and Gilovich (2010) asked respondents to rank the income-happiness correlation among a set of other empirical correlations.

[5]LLMs have also been used for other purposes, such as experimental interventions. For example, Costello et al. (2024) use an LLM to persuade respondents through dialogue.

about a real decision they are facing and then uses that qualitative information, in real time, to construct a closed-ended survey question tailored to the respondent. Our approach goes beyond fixed survey instruments, combining the flexibility of qualitative interviews with the discipline of closed-ended measurement.

The rest of the paper proceeds as follows. Section 2 presents the conceptual framework. Section 3 describes the research design. Section 4 discusses implementation details and measurement. Section 5 presents the main results. The last section concludes.

2 Conceptual Framework

This section presents a simple conceptual framework that motivates and informs the experimental design.

2.1 Income and Satisfaction

Let i index individuals and t index time periods. Let $S_{i,t}$ denote life satisfaction, which we also refer to simply as satisfaction, and let $I_{i,t}$ denote the individual's log income. The model centers on the marginal satisfaction from income (MSI), defined as the long-run effect on life satisfaction per unit increase in log income. More precisely, there are two objects of interest:

- Individual i's own long-run marginal satisfaction from log income, given by β_i. We refer to this as self-MSI.
- The population-wide average marginal satisfaction from log income, given by $\bar{\beta} > 0$. We refer to this as average-MSI.

Individuals share a prior over average-MSI,

$$\bar{\beta} \sim \mathcal{N}(\bar{\beta}_0, \kappa_0^{-1}), \qquad \kappa_0 > 0, \tag{1}$$

Conditional on true average-MSI $\bar{\beta} > 0$, the individual's self-MSI is drawn from the population:

$$\beta_i \mid \bar{\beta} \sim \mathcal{N}(\bar{\beta}, \sigma_\beta^2), \qquad \sigma_\beta^2 > 0. \tag{2}$$

This relationship links beliefs about self-MSI and average-MSI.

The model assumes a log-linear relationship between income and life satisfaction, a functional form with strong empirical support in the literature (e.g., Stevenson and Wolfers, 2013). This specification simplifies the model: individuals may differ in their MSI, but each individual's MSI can be summarized by a single number. It also simplifies the connection

between self-MSI and average-MSI in the learning model: individuals can use their own experiences to learn about others, and others' experiences to learn about themselves, without separately accounting for the income levels at which those experiences occurred.

2.2 Income Adaptation

The key feature of the true satisfaction process is that satisfaction depends not only on current log income ($I_{i,t}$) but also on the gap between current log income and the reference log income ($I_{i,t}^{\text{ref}}$):

$$S_{i,t} = \alpha_i + \beta_i I_{i,t} + \theta\beta_i(I_{i,t} - I_{i,t}^{\text{ref}}) + \varepsilon_{i,t}, \qquad \theta > 0, \tag{3}$$

where $\varepsilon_{i,t}$ is a mean-zero shock and where reference income gradually adjusts toward past income:

$$I_{i,t}^{\text{ref}} = (1-\lambda)I_{i,t-1}^{\text{ref}} + \lambda I_{i,t-1}, \qquad \lambda \in (0,1], \tag{4}$$

where λ governs the speed of adjustment. We assume that changes in log income are exogenous.[6] As a result, a permanent one-unit increase in log income initially raises satisfaction by $(1+\theta)\beta_i$. As reference income adjusts, however, a fraction $\theta/(1+\theta)$ of that gain fades, leaving a lasting gain of β_i.[7]

This adaptive reference income is intended to capture several phenomena that have been widely discussed in the psychology and economics literature, some of which are mentioned explicitly on the last screen of the information treatment. For example, the reference point could partly reflect income aspirations: as individuals become richer and adjust to their new circumstances, what they consider a sufficient income rises along with their actual income. A closely related phenomenon is income comparisons: as individuals become richer, they tend to interact with richer people, and these new comparisons may raise their income aspirations (Clark and Senik, 2010). The reference point could also reflect hedonic adaptation. Individuals adapt to repeated stimuli, so while a new experience, such as driving a new car, may feel especially rewarding at first, the feeling fades with repeated exposure until driving the new car feels much like driving the old one. Hedonic adaptation is supported by evolutionary explanations and neuroscience findings (Rayo and Becker, 2007) and has also been documented

[6] For tractability, innovations in log income are Gaussian. Appendix A.1 specifies the random walk and independence conditions and establishes the exact signal means used in the predictions.

[7] The model defines the long run as the point at which reference income has fully adjusted, without assigning that adjustment a fixed calendar horizon. In the empirical application, the short run refers to relatively immediate satisfaction responses—for example, how it feels to buy a new car during the first week—whereas the long run refers to effects that remain after there has been time to adapt, such as whether the purchase continues to affect life satisfaction 12 months later or beyond.

using life-satisfaction data (e.g., Di Tella et al., 2010; Galiani et al., 2018).

2.3 Misspecified Bayesian Learning

Individuals learn about income's lasting effect on satisfaction from changes in their own lives and in the lives of others. Levels of satisfaction are hard to interpret: in any given year, income may change alongside many other events that affect health, relationships, family, leisure, and other dimensions of life. Short-run changes are often more salient and easier to connect to particular events. For example, someone may remember feeling especially satisfied after receiving a large bonus or during an expensive vacation. From such episodes, they may infer that income raises satisfaction. To formalize how individuals extract information from these experiences, we assume they estimate a regression of changes in their own satisfaction ($\Delta S_{i,t}$) on changes in their own log income ($\Delta I_{i,t}$). They do the same with the changes in satisfaction and log income they observe among their peers.[8] As a result, the individual neglects adaptation by interpreting the first-difference estimates as long-run effects of income on satisfaction. In reality, these estimates capture short-run responses and therefore overstate the long-run effects on average.

Because β_i is drawn from the population distribution, learning is hierarchical: own data are directly informative about self-MSI but also informative about average-MSI, while peer data are directly informative about average-MSI but also informative about self-MSI. Beliefs are therefore biased because Bayesian updating uses a misspecified model of how income affects satisfaction over time. To isolate adaptation neglect, we assume that individuals share a correctly centered prior mean $\bar{\beta}_0 = \bar{\beta}$ for average-MSI. They combine this prior with own and peer evidence to form a belief about average-MSI, denoted $\tilde{\bar{\beta}}$. Their belief about self-MSI, denoted $\tilde{\beta}_i$, combines their own-data estimate with this inferred population mean. For tractability, subjective signal distributions are Gaussian—Appendix A.2 states the assumptions and gives the updating formulas. As a result, both beliefs are weighted averages of the prior mean and the two sources of evidence.

2.4 Predictions

The model delivers three predictions that guide the empirical analysis. The first prediction relates to the mean biases in beliefs:

[8] The regressions are a metaphor; we do not assume that individuals literally estimate them. Appendix A.1 defines the estimators and establishes their exact means.

Prediction 1. On average, individuals overestimate both self-MSI and average-MSI, with weakly stronger overestimation of self-MSI.

$$\mathbb{E}\Big[\tilde{\beta}_i - \beta_i\Big] \geq \mathbb{E}\Big[\tilde{\bar{\beta}} - \bar{\beta}\Big] > 0. \tag{5}$$

The first inequality is strict when own experience is informative.

Both beliefs are upward biased on average because individuals neglect adaptation and overestimate how much of the initial satisfaction gain from higher income will persist. The larger mean bias in self-MSI beliefs arises because personal experience carries more weight when individuals assess their own MSI than when they assess average-MSI. An income gain provides direct evidence about one's own satisfaction, but only indirect evidence about the average person's, since people differ in how income affects their satisfaction.

The second prediction concerns how individuals would update their beliefs in response to information. Let m denote a signal about true average-MSI, and let $\tilde{\beta}_i(m)$ and $\tilde{\bar{\beta}}(m)$ denote posterior mean beliefs after observing it, holding fixed the individual's own and peer data. We consider the benchmark realization $m = \bar{\beta}$, which reports the true average. The individual treats the signal as informative while allowing for uncertainty about its accuracy.[9]

Prediction 2. On average, accurate information lowers both average-MSI and self-MSI beliefs, with a weakly larger reduction in average-MSI beliefs.

$$\mathbb{E}\Big[\tilde{\bar{\beta}}(\bar{\beta}) - \tilde{\bar{\beta}}\Big] \leq \mathbb{E}\Big[\tilde{\beta}_i(\bar{\beta}) - \tilde{\beta}_i\Big] < 0. \tag{6}$$

The first inequality is strict when own experience is informative.

The intuition is that adaptation neglect makes average-MSI beliefs too high, so accurate information moves them downward. Self-MSI beliefs also fall because they partly reflect the inferred population mean. Own experience continues to anchor self-MSI beliefs, making their mean reduction smaller when own experience is informative.

The third prediction concerns how the provision of information affects choices. Let a_i denote a normalized income-generating action, representing choices such as hours worked, effort to obtain a higher-paying job, or pursuit of a higher-paying career. For tractability, assume that log income is linear in this action, with a common slope normalized to one. The individual values life satisfaction and other benefits of income and faces quadratic action costs. Up to terms independent of a_i, utility is

$$V_i(a_i; m) = \Big[\tilde{\beta}_i(m) + \gamma_i\Big]a_i - \frac{c}{2}a_i^2, \tag{7}$$

[9] Appendix A.4 specifies this signal structure.

where $\gamma_i > 0$ captures other benefits of income and $c > 0$ governs the curvature of action costs and is common across individuals. Starting from an interior-choice benchmark, let Δa_i denote the post-information minus pre-information change in the income-generating action.

Prediction 3. On average, accurate information lowers the income-generating action.

$$\mathbb{E}[\Delta a_i] < 0. \tag{8}$$

The intuition is that accurate information lowers self-MSI beliefs on average. This lowers the perceived benefit of income-generating actions, while leaving their direct cost unchanged, so individuals choose less of the income-generating action on average.

This framework focuses on income, but adaptation neglect need not be confined to material gains. The relevant assumption is that individuals overestimate the lasting satisfaction gains from material improvements relative to those from non-material improvements, such as more time with family, a shorter commute, or more sleep. For example, driving a new car off the dealership lot may produce an initial "high" that fades as the novelty wears off (Schwarz and Xu, 2011). By comparison, the benefits of sleep, time with family, or engaging activities may be renewed with each experience and depend less on novelty (Hsee et al., 2009; Frey and Stutzer, 2014).

3 Research Design

3.1 Overview

A sample survey instrument for the baseline survey is provided in Appendix E, and the corresponding instrument for the follow-up survey is provided in Appendix F.[10] Both surveys are described in more detail below. The baseline survey began with questions used to construct job-choice scenarios tailored to each respondent and with an AI-led interview used to elicit a real-world income-related decision. We then elicited prior self-MSI and average-MSI beliefs. Respondents were randomly assigned either to the information treatment or to a control condition. After this information-treatment stage, we re-elicited the same beliefs and measured choices involving trade-offs between income and non-income considerations. The follow-up survey, conducted one month later, re-elicited the main belief measures, repeated the choice-framing job-choice measures, and asked respondents about the real-world decision discussed at baseline.

[10]The actual survey varied across respondents because some questions depended on earlier answers and because key elements were randomized, including treatment assignment and framing order.

3.2 Prior Average-MSI and Self-MSI Beliefs

Before any information was provided, we elicited respondents' prior self-MSI and average-MSI beliefs. More precisely, we elicited the perceived effect of a 20% income increase on life satisfaction on a 0–100 scale. Respondents may be largely unfamiliar with reporting life satisfaction on such a scale, so they may not have an intuitive sense of how large a one-point satisfaction increase is. This is analogous to asking respondents to evaluate prices in a currency different from the one they are most accustomed to. While we cannot eliminate this type of measurement error entirely, we can at least take steps to mitigate it. The first step was to ask them to report their overall life satisfaction using the same 0–100 scale, so they had the opportunity to familiarize themselves with the concept and the scale. The second step takes inspiration from the analogy with a foreign currency: even if the exchange rate is known, having some benchmark prices may help respondents get an intuitive sense of the scale of the foreign currency. To that end, immediately before the prior-belief elicitation, we showed respondents some benchmarks for the effects of non-income changes on life satisfaction, based on scientific evidence: the death of a spouse is associated with a 7.9-point reduction in life satisfaction in the year following the event; becoming unemployed is associated with a 5.2-point reduction; getting married is associated with a 4.6-point increase; and deactivating Facebook for four weeks increases life satisfaction by about 3.1 points during that period.[11]

We elicited two types of prior beliefs: average-MSI beliefs and self-MSI beliefs. For average-MSI beliefs, respondents considered an average person who earns $50,000 per year and reports life satisfaction of 75 out of 100. The question then asked how much an unexpected 20% income increase, corresponding to an additional $10,000 in income, would change that person's life satisfaction over the next 12 months. For self-MSI beliefs, respondents were asked the analogous question about themselves: how much an unexpected 20% increase in their own income would change their own life satisfaction over the next 12 months. In both cases, respondents answered using the following options: a decrease of 1 point or more, no change, an increase of 1 point, an increase of 2 points, continuing in one-point increments up to an increase of 14 points, and an increase of 15 points or more. To keep the baseline survey short, we elicited confidence in these beliefs only in the shorter follow-up survey.

The log-linear specification substantially simplifies belief elicitation. Under this specification, a single question about the satisfaction gain from a 20% income increase suffices to characterize each MSI belief, and the response can be extrapolated to other starting incomes. Without a functional-form restriction, characterizing these beliefs would require eliciting the perceived income-satisfaction relationship across a range of starting incomes. One justifica-

[11] The source for the first three effects is Clark and Georgellis (2013), and the source for the last one is Allcott et al. (2020).

tion for this functional form is the strong empirical support for a log-linear specification of the actual relationship (e.g., Stevenson and Wolfers, 2013). A separate question is whether individuals perceive this relationship as log-linear or instead believe it to be more or less concave. Our data suggest that, although respondents may perceive slightly greater concavity than the logarithmic specification implies, the log-linear assumption provides a reasonable approximation—see Appendix D.2 for details.

A central mechanism through which income can raise life satisfaction is the additional consumption it enables. Our framework and empirical approach could therefore also be applied to the marginal satisfaction from *consumption* instead of marginal satisfaction from *income*. We focus on income for practical reasons. The main reason is that the more extensive literature on income and life satisfaction provides a broader evidence base for constructing the scientific benchmark. Income is also generally easier to measure than consumption in a survey.

3.3 Information Treatment

After prior beliefs were elicited, subjects were randomly assigned to the treatment arm or the control arm, each with a 50% probability. Following best practices in information-provision experiments, we made the randomization explicit: respondents were told that some subjects would be randomly selected to receive the information treatment and that they would learn on the next screen whether they had been selected.[12] Subjects assigned to the treatment arm received a short information module summarizing the scientific evidence on the effect of income on life satisfaction. The central message of this module was that the effect of a 20% increase in income on life satisfaction is positive but modest, approximately one point on a 0–100 life-satisfaction scale. Respondents in the control arm saw an otherwise identical survey without the information module.

The information module was split across three consecutive screens, reproduced in Figure 1. This structure allowed respondents to digest the evidence gradually. The first screen focused on the empirical correlation between income and life satisfaction. It presented evidence from the 2018 General Social Survey, a large U.S. survey that asked respondents about both household income and life satisfaction. An explainer video showed step by step that a 20% increase in income is associated with an increase in life satisfaction of approximately

[12]This explicit randomization is meant to prevent subjects from making any unnecessary inferences from seeing the information module: e.g., respondents could think that they were shown the module because their prior beliefs were inaccurate.

one point.[13] For more details about the explainer video, including screenshots and the full narration, see Appendix B.2. This first screen was meant to make the magnitude transparent, but it was framed explicitly as correlational evidence. We chose the 2018 wave of the GSS purely for data availability and convenience, but the magnitude of this association is consistent with estimates that use different surveys. For example, using Gallup data on more than 450,000 U.S. respondents, Kahneman and Deaton (2010) find that life evaluation (the 0–10 Cantril ladder, an evaluative measure akin to life satisfaction) rises steadily with log income. When converted to the same 0–100 scale as ours, their estimates imply that a 20% increase in income is associated with a one-point increase in well-being—the same magnitude as our benchmark.[14] Moreover, the association between income and life satisfaction turned out to be similar in magnitude when estimated with the survey responses from our experiment; see Appendix C.4 for details.

The goal of the second screen was to show that, while the correlations reported in the first screen do not imply a causal effect, there is causal evidence that supports the same finding. It discussed Lindqvist et al. (2020), who use the random allocation of lottery prizes as a natural experiment to study the long-run effects of wealth on psychological well-being. The module explained that the study followed lottery winners for many years after the lottery event and found persistent effects on overall life satisfaction. A lottery windfall can finance additional consumption over many years. To make these wealth gains comparable to sustained increases in annual income, we use the annuitized estimates reported in Table A10 of the Online Appendix to Lindqvist et al. (2020). The authors convert each prize into an equivalent annual payment over 20 years, assuming a 2% real interest rate. Based on their estimates, the implied effect of a 20% increase in annual income is about 0.7 points on a 0–100 life-satisfaction scale. We round this estimate up to 1 point.

The findings of Lindqvist et al. (2020) are consistent with randomized cash-transfer experiments showing that additional financial resources can improve subjective well-being, including Haushofer and Shapiro (2016) in Kenya, Dwyer and Dunn (2022) across seven countries, and Vivalt et al. (2026) in the United States. The last study is closest to our setting, making a comparison of magnitudes particularly informative. Because our MSI elicitation asks about satisfaction during the first 12 months following an income increase, we focus on their

[13] This analysis is based on the raw correlation, but including control variables would have produced the same +1 feedback; see Appendix C.4 for details. Although we would have preferred to include control variables, the video presents the raw correlation to keep it simple for a non-technical audience.

[14] Kahneman and Deaton (2010) report in Table 2 (second row, fourth column) that mean life-satisfaction scores, measured on a 0–10 scale, differ by 0.23 between the $7,500–$9,999 and $5,000–$7,499 monthly income brackets. The annualized midpoints of these brackets are approximately $105,000 and $75,000, respectively, implying a 40% difference in income. Assuming, as they document, that life evaluation is log-linear in income, the 20% income increase corresponds to an increase of 0.125 ($= 0.23 \cdot \ln(1.2)/\ln(1.4)$) points on the 0–10 scale, or equivalently 1.25 points on a 0–100 scale.

first-year effects.[15] A rough conversion of their estimates implies that a 20% income increase raises life satisfaction by approximately 1.3 points on a 0–100 scale, close to our scientific benchmark of one point.[16]

The third screen focused on causal mechanisms, that is, the scientific evidence that may help explain why MSI is lower than some respondents expect. It discussed the phenomenon of hedonic adaptation, with a hyperlink to the source study Di Tella et al. (2010): people may initially feel better after an income gain, but part of that effect can fade as they adjust to the new circumstances. This screen also discussed the role of rising aspirations, with a hyperlink to the source study Stutzer (2004): as income increases, the amount people view as sufficient may also rise. Finally, the module noted that while income can relieve financial constraints, many other determinants of life satisfaction—such as health, relationships, and working conditions—remain important even as income rises.

The information treatment is designed to provide the empirical counterpart of the accurate signal in the conceptual framework. It provides information about average-MSI, not a personalized estimate of each respondent's self-MSI. The information module cited the underlying research studies and used neutral language: it summarized evidence on the income-satisfaction relationship without recommending any particular belief, choice, or lifestyle. In many information-provision experiments, the information treatment can be reduced to a single familiar statistic. For example, in studies of inflation expectations, respondents are typically accustomed to seeing official inflation statistics. In that context, the information treatment may simply provide the official inflation rate over the previous year (Cavallo et al., 2017). Our setting is different. Most respondents are unlikely to have encountered scientific estimates of the income-satisfaction relationship before, and a bare statement that a 20% income increase raises life satisfaction by about one point may therefore be difficult to interpret or trust. For this reason, the information module presented not only the headline estimate but also the evidence behind it. The goal was to make the signal credible and interpretable while keeping the message neutral.

[15] Consistent with adaptation, the gains diminish after the first year and are no longer statistically significant by the second year, even though transfers continue.

[16] Figure IX shows a first-year effect of approximately 0.23 points on a 0–10 life-satisfaction scale, equivalent to 2.3 points on our 0–100 scale. Participants received $12,000 annually in the treatment group and $600 in the control group. Adding these transfers to average baseline household income of approximately $29,900 gives $41,900 and $30,500, respectively. Assuming a log-linear relationship, rescaling to a 20% income increase gives $2.3 \times \ln(1.2)/\ln(41{,}900/30{,}500) \approx 1.3$. This approximate calculation uses baseline income and gross transfers, without adjusting for treatment-induced changes in earnings.

3.4 Posterior Beliefs

Immediately after the information-treatment stage, we re-elicited self-MSI and average-MSI beliefs. Before doing so, the survey clarified that all respondents were given the opportunity to reassess their answers, regardless of their prior responses and regardless of whether they had received the information treatment. This clarification was intended to reduce the concern that respondents would interpret the repeated question as a signal that their initial answer had been incorrect. Comparing posterior beliefs between the treatment and control arms allows us to test whether information about average-MSI shifts both average-MSI and self-MSI beliefs. In the follow-up survey, we elicited these two beliefs again, to measure whether any effects of the information on beliefs persisted one month later.

3.5 Job-Choice Scenarios

After eliciting posterior beliefs, we measured a series of post-treatment outcomes, focusing on job-choice scenarios and real-world decisions involving trade-offs between income and non-income considerations. To tailor the job-choice scenarios to each respondent, we asked four questions near the beginning of the survey about the respondent's current or most recent job: occupation or job title, annual earnings, typical work hours, and typical commute length.[17] We used this information to construct job-choice scenarios that would be realistic for each respondent. More precisely, each respondent faced three job-choice scenarios. In each scenario, one option was similar[18] to the respondent's current or recent job, while the other option paid more but required sacrificing a non-income benefit. The three trade-offs were higher income versus a longer commute, higher income versus longer working hours, and higher income versus fewer hours of sleep.

It is easiest to explain these scenarios using one concrete example. Panel A of Figure 2 shows the scenario corresponding to the trade-off between income and hours worked. Respondents saw a table with a side-by-side comparison between two job offers: both had the same position title, but relative to the offer on the left, the offer on the right had 20% higher earnings but 20% more work hours. Respondents had to choose one of four responses: "definitely choose A", "possibly choose A", "possibly choose B", or "definitely choose B". The other two job-choice scenarios use the same format. In the income-versus-commute scenario, one job offered 15% higher earnings but had an 80% longer commute. In the income-versus-sleep scenario, one job offered 30% higher earnings but reduced sleep by 25%, from 8 hours

[17] Respondents who reported not currently working were asked about their most recent job instead.

[18] We say "similar" instead of "identical" because we rounded the default option's wages to the nearest dollar and commute time to the nearest hour, and the sleep-time numbers are fully hypothetical since we did not ask how many hours individuals sleep.

to 6 hours.[19]

For each scenario, respondents answered two sets of questions, in randomized order. The main set (the choice framing) asked which of the two jobs the respondent would choose in each of the three scenarios. The alternative set (the satisfaction framing) asked which option would make the respondent more satisfied with their life overall. The use of both framings allows us to separately identify the effects of the information on utility-maximizing and satisfaction-maximizing behavior (Becker and Rayo, 2008; Benjamin et al., 2014). In the follow-up survey, we repeated the three job-choice scenarios, but for the sake of brevity, we did so only for the choice framing.

3.6 Real-World Decisions

Respondents were asked about a decision that they were currently facing or would soon face, in which they were weighing an income consideration against a non-income consideration. Respondents were asked to describe that decision in their own words. They then participated in a short AI-led interview conducted through an LLM called in real time from within the survey. The LLM was prompted to ask questions to clarify the relevant alternatives and the trade-off involved. For more details about how the AI-led interview was conducted, see Appendix B.3. At the end of the interview, the AI interviewer generated a personalized binary choice and shared it with the respondent. More precisely, the AI interviewer showed a description of Option A, a description of Option B, and a description of the underlying trade-off. Respondents were then asked to confirm whether these descriptions accurately reflected the real-world decision they were facing. We also asked them when they expected to make this decision.

Most importantly, we asked respondents which of the two options, Option A or Option B, they were more inclined to choose. For this elicitation, we closely followed the same style used for the job-choice scenarios. More precisely, we created a table that summarized all the information relevant to the decision. An example of this table is shown in Panel B of Figure 2. This respondent was deciding whether to work extra hours (Option A) or not (Option B). The trade-off was described as making some extra money for their wedding versus having less time with their family and risking burnout. In the main choice framing, subjects selected one of four responses: "definitely choose A", "possibly choose A", "possibly choose B", or "definitely choose B". As in the job-choice module, we also asked the same question but under the satisfaction framing, in randomized order. The follow-up survey included a few

[19] We rounded commute times to the nearest hour and imposed minimums of 3 weekly commuting hours and 5 weekly work hours for the alternative job. The commuting minimum applied to respondents who reported a daily commute under 20 minutes (44.4%, including the 26.9% who reported no commute).

questions related to this real-world decision. Respondents were asked whether they had made the real-world decision discussed in the AI-led interview and, if so, what the outcome was. If they had not made a decision yet, we asked them whether the two options were still available and, if so, which of the two options they intended to choose in the future (see Appendix C.6 for how we combine these answers into a single outcome).

3.7 Additional Outcomes

The surveys included a few additional questions. We pre-registered an additional outcome related to a real-world behavior: the number of paid surveys the respondent completed on the Prolific platform. Because participants complete surveys for pay on the Prolific platform, completing more surveys allows them to earn more income. Indeed, in the real-world decision module, some participants reported that completing more Prolific surveys was the income-related decision they were considering. We also included three secondary outcomes related to income aspirations to explore whether the treatment effects operated through an aspirations channel. For brevity, we discuss these three questions and their results in Appendix D.11.

4 Implementation Details

4.1 Fielding the Baseline Survey

The experiment was pre-registered at the AEA RCT Registry (AEARCTR-0018251) and was reviewed and approved in advance by the Institutional Review Board at UCLA. We recruited respondents for the baseline survey between April 3 and April 9, 2026. Subjects were recruited via Prolific. We advertised the study as "Scientific study". We limited eligibility to U.S. residents and required respondents to complete the survey on a laptop or desktop computer. 3,002 respondents completed the baseline survey.[20] The median completion time was 18.3 minutes.[21] About 87.8% of respondents reported that the survey was very easy or somewhat easy.

4.2 AI-Use and Data Quality Checks

The baseline survey included several measures to validate the quality of the survey data. The Prolific platform has its own measures in place to detect and prevent AI use in surveys (Gordon, 2026). According to independent accounts, at the time that we fielded the survey

[20] Some respondents started the baseline survey but did not complete it and therefore are not part of our main analysis sample—for more details, see Appendix D.9.

[21] The baseline survey had a participation reward of $3.

the estimated share of Prolific participants using AI was reported to be very small (e.g., Çelebi et al., 2026; Affonso, 2026). However, in order to be extra cautious, we followed best practices by including our own independent checks, as in Çelebi et al. (2026). The main video-code AI-use check is based on showing a short video and asking respondents to enter the numbers shown in the video. The video-code check was passed by 95.0% of subjects, so the 5% who failed the check provide an upper bound for the potential share of AI-assisted responses, as humans could fail this check if they are not paying close attention. We also included CAPTCHA verification, and the mean reCAPTCHA score was 0.972 on a 0–1 scale, where higher values indicate a greater likelihood that the respondent is human. We also included a couple of standard attention checks at the end of the survey, which were passed by 98.0% of subjects. As another data quality check, the distribution of completion times suggests there is no concern about substantial speed-taking: only 0.93% of respondents completed the survey in under seven minutes.

4.3 Sample Restrictions

We start by excluding the 106 respondents (3.53%) who clearly failed a data-quality check: those who failed at least one of the two attention checks at the end of the survey, and those whose answer to the video-code check does not resemble the code shown in the video (e.g., a blank answer, text, or a wrong set of digits). Additionally, in information-provision experiments, it is important to handle outlier beliefs appropriately. Some respondents may report wildly inaccurate guesses not because they genuinely hold such extreme beliefs, but because they misunderstood the question or were not paying sufficient attention.[22] If respondents are confused about the basic concept being measured, it is hard to interpret what the effects of information may be. To reduce sensitivity to outliers, we follow standard practice in information-provision experiments and exclude respondents with the most extreme prior beliefs or misperceptions (e.g., Fuster et al., 2022; Cullen and Perez-Truglia, 2022; Giaccobasso et al., 2025). Because prior beliefs in our study are measured discretely, we drop respondents who selected one of the two most extreme response options: "Decrease life satisfaction by 1 point or more" and "Increase life satisfaction by 15 points or more".[23] In the baseline specification, we take a conservative approach and exclude outliers in prior average-MSI be-

[22]For example, if a respondent reports a recent annual inflation rate of 50% in a country where the true rate is 5%, it is likely that the respondent misunderstood the question or lacks a basic understanding of percentages.

[23]Throughout the analysis, we code the two endpoint categories as -1 and +15, respectively. Thus, the belief measures are effectively bottom- and top-coded at -1 and +15 satisfaction points.

liefs.[24] Importantly, since prior beliefs were elicited before treatment assignment, excluding respondents on the basis of their priors does not compromise the validity of the experimental variation. After applying these restrictions, the main analysis sample at baseline contains 2,775 respondents. We apply the same video-code rule to the follow-up responses, which excludes a further 21 respondents and leaves us with 2,264 follow-up observations. We also show robustness checks with alternative sample restrictions and without the prior-belief restriction altogether.

4.4 Descriptive Statistics and Randomization Balance

Column (1) of Table 1 provides summary statistics for the sample: 48.8% of respondents are female, 64.4% are White, 75.8% are employed, average household income is approximately $91,010, and average baseline life satisfaction is 62.1 on the 0–100 scale. Columns (2) and (3) of Table 1 provide a balance test. These columns compare average pre-treatment characteristics in the control and treatment arms, respectively. The characteristics include a range of demographic variables—such as age, gender, and education—as well as prior beliefs about the average-MSI and self-MSI. Column (4) reports the p-value from a two-sided test of equality of means, with standard errors shown in parentheses. None of the differences is statistically significant, and the joint test of equality across all variables has a p-value of 0.906. As is common in online samples, our subject pool is somewhat different from the general U.S. population: respondents are somewhat younger and more educated—see Appendix C.1 for more details.

4.5 Job-Choice Scenarios

The job-choice module used respondents' own employment information to construct job-choice scenarios. These scenarios could only be constructed for respondents who had current or past employment and entered valid information, a group that includes almost all respondents (98.1%).[25] On average, subjects earned $60,438 annually, worked 35.8 hours weekly, and commuted to work for 31.4 minutes daily.

The outcomes are coded so that higher values correspond to greater willingness to choose the higher-income option. For example, in the income-versus-work-hours scenario, Offer B

[24] For respondents who do not report an extreme average-MSI belief but do report an extreme self-MSI belief, our interpretation is that they understood the question but believe they themselves have extreme preferences.

[25] In the main analysis sample, 75.8% of respondents currently have a job, 22.2% do not currently have a job but had one before, and 1.95% have never had a job. For the first group the job questions refer to the current job, and for the second group to the most recent job.

was the one that offered higher earnings but longer hours. Thus, the dependent variable is coded as follows: 1 if the respondent chooses "definitely choose A", 2 if "possibly choose A", 3 if "possibly choose B", and 4 if "definitely choose B". We coded the other two job-choice scenarios accordingly. We also constructed a job-choice index that averages across the three individual outcomes.

In the work-hours and sleep scenarios, the two offers were reasonably balanced in relative attractiveness on average. In the income-versus-work-hours scenario, the average outcome in the control group is 2.45. Since this value is close to 2.5 (the middle of the 1–4 scale), it means that both offers were about equally attractive on average. For the income-versus-commute scenario, the average outcome is 1.92, meaning that the offer with higher income was, on average, clearly less attractive. For the income-versus-sleep scenario, the average outcome is 2.81, meaning that the offer with higher income was, on average, a bit more attractive.[26]

4.6 Real-World Decision Module

A manual inspection of conversations between respondents and the AI interviewer suggests that this module went smoothly. For verbatim extracts from the full interview transcripts, see Appendix B.4. Additionally, in the feedback form at the end of the survey, some subjects explicitly mentioned that they enjoyed the conversation with the AI interviewer. Overall, the AI interviewer was able to produce a description of the respondent's decision for the vast majority (98.4%) of respondents.[27] When shown this description of their decision, 88.4% of respondents rated it as accurate, 10.7% rated it as somewhat accurate, and the remaining 0.90% rated it as inaccurate.

The real-world decisions varied widely from respondent to respondent. We provide a more detailed analysis below, but we can start with a few interesting examples. For example, some decisions concerned personal consumption or tastes: one respondent was weighing whether to work longer hours to pay for an expensive wedding, and another was choosing between a higher-paying career in law and the pursuit of a passion for the circus. In other decisions, serious health issues were involved—for example, a respondent was considering delaying cancer treatment to avoid a loss of income. Finally, other individuals preferred to discuss family decisions: one was deciding whether to have a second child, knowing that it would affect her earnings, and another was contemplating leaving an unhappy marriage despite the financial hardship that would likely follow.

[26]These averages correspond to the (main) choice framing. The averages are slightly smaller for the satisfaction framing—see the bottom of Table 2 for more details.

[27]The AI interviewer was prompted to return an exit code in case the respondent refused to engage in a conversation. Additionally, for a few observations the AI was unable to produce any valid output, which could be due to a failed connection to the internet or the server.

Constructing the outcome variable for the real-world decisions is more challenging. For the job-choice scenarios, the outcome variable is straightforward: because we constructed the scenarios, we know exactly which of the two offers has higher earnings. For the real-world decisions, however, the choice options were generated in real time by the AI interviewer. As a result, we need to determine ex post which of the two options offers higher income. For brevity, throughout the rest of the document we use terms such as "income considerations" and "higher-income", but these terms should be understood broadly. In most cases, they refer to choices that directly involve higher income, such as choosing a higher-paying job or increasing income by working additional hours. In other cases, however, they refer to choices that improve the respondent's financial situation without necessarily increasing income, such as moving to a state that has a lower cost of living but is far from family, or choosing a less pleasant job with more stable income.

Consider the previous example where Option A is working extra hours to pay for the wedding, Option B is not working extra hours, and working the extra hours would provide more income but leave less time to spend with family and also risk burnout. In this case, we observe a clear trade-off between income and non-income considerations, and it is also clear that Option A is the one that offers higher income. As a result, for this respondent the dependent variable takes the value 1 if the respondent chooses "definitely choose B", 2 if "possibly choose B", 3 if "possibly choose A", and 4 if "definitely choose A". However, for other subjects, Option B may be the one that offers higher income. For some decisions, it may be ambiguous which of the two options offers higher income, or there may not be a clear trade-off to begin with.

We used an LLM to evaluate respondents' decisions observation by observation, supplemented by human coding for the most difficult cases. Appendix C.6 describes the prompt and coding scheme, and the full prompt is reproduced in the replication package; below we provide a brief summary. For each respondent, the LLM was asked to identify which of the two options offered higher income and to indicate whether it was highly certain in that classification. The LLM could also indicate that neither option had a clear income dimension, or that it was ambiguous which option offered higher income. In addition, the LLM rated the difficulty of the classification. For the 211 most difficult cases among all baseline completers, a human coder reassessed the LLM's classifications.

In the baseline specification, we apply the following exclusion criteria. We start by dropping the 1.62% of observations for which the AI interviewer did not produce valid output and exited. The percentages for the remaining criteria are each expressed as a share of the decisions *with* valid output. First, among these, we keep only the observations for which the respondent rated the AI output as accurate, by dropping those rated as somewhat accurate

(10.7%) or inaccurate (0.90%). We also drop the decisions for which neither option had a clear income dimension, or for which it was ambiguous which of the two options offered higher income (4.10%). Lastly, we drop the decisions where a trade-off was identified only with low confidence by the LLM (25.8%). The combination of these exclusion criteria means that the number of observations for the real-world outcomes (1,711) is significantly lower than the corresponding number for the job-choice scenarios (2,721), or 37.1% lower.[28] This smaller sample size implies that the results for real-world decisions are estimated with lower precision. We also report sensitivity analyses using alternative exclusion criteria.

LLM classifiers are not fully reproducible: the same prompt can return a different label each time it is run. We therefore compare our LLM classification against another one that does not make any LLM call and is completely deterministic. To do so, we take the 1,711 observations from the baseline specification and identify which of the two options offers higher income using a rule-based keyword classifier. At classification time, this classifier makes no LLM calls and does not use any labels produced by the LLM classifier. Instead, it applies a fixed dictionary of 255 income-direction phrases to Option A, plus two shorter lists of phrases to the description of the trade-off: phrases such as "ask for a raise," "work more hours," or "accept the promotion to manager" mark the option they describe as the higher-income one, and the decision is left unclassified when no cue matches. Because we developed the dictionary with LLM assistance while inspecting our data, this exercise is a reproducibility check rather than an independent validation of the LLM classifications. Among the 1,711 baseline decisions, both the keyword classifier and the LLM classifier assign an income direction to 1,471 of them, and on those they agree on which option offers higher income in 97.1% of cases; see Appendix C.6 for more details. As a final check, we re-ran the same LLM classifier with the exact same prompt, inputs, and model. Compared to the original classification, the re-run assigns the exact same code in 94.1% of cases, and, among the 1,711 decisions used in our baseline analysis, the re-run would reverse the income direction of only 0.29%.

Next, we provide a descriptive analysis of the real-world decisions subjects were facing. In terms of timing, the typical decision horizon was about one month. More precisely, when asked one month later in the follow-up survey, 71.5% of subjects reported that they had already made a decision.[29] The real-world decisions appear to be fairly balanced: at baseline, the average outcome in the control group is 2.47. Since this value is close to 2.5,

[28]The percentages reported above do not add up to 37.1% because the exclusion criteria are not mutually exclusive; for example, a decision rated as inaccurate may also have an ambiguous income direction.

[29]In comparison, in the baseline survey 33.1% of respondents expected to make a decision within one month. The discrepancy is likely due to the fact that many of these were not one-off decisions but recurring ones: among respondents who had not yet reached a decision at follow-up, 93.0% reported that the two options were still available.

the midpoint of the 1–4 scale, this means that the higher-income and lower-income options were, on average, about equally attractive.[30] Between the baseline and follow-up surveys, many things can happen that may sway subjects in one direction or the other. Thus, while we do not expect choices to be identical between baseline and follow-up, it is still useful to compare them. Among respondents who had reached a decision at follow-up, the option they preferred at baseline coincides with what they ultimately chose one month later in 75.0% of cases.

Next, to provide descriptive statistics about the types of income and non-income considerations subjects were trading off, we use BERTopic, a fully unsupervised topic model (Grootendorst, 2022), in combination with LLM-written descriptions of each decision. Appendix C.5 describes the full procedure and provides additional results. The results are presented in Figure 3. Panel A summarizes the income considerations. The two largest categories are similar in size and together account for more than half of respondents: 27.5% are considering working more—for instance through extra hours, additional shifts, or a second job—and 27.1% are considering a higher-paying job. The remaining categories are smaller and more dispersed, including relocating for higher pay, freelance or gig work, returning to work, taking time off, working from home, promotions, and retirement-related decisions. Panel B summarizes the non-income considerations. The most common category is family time, which accounts for 30.6% of respondents. The second-largest category, at 21.1%, is leisure: free time, rest, hobbies, and work-life balance. The remaining categories are again smaller and dispersed and include schedule flexibility, proximity to family and friends, mental health, remote work, and shorter commutes. Panel C reports the most common pairs of income and non-income categories. The most frequent trade-off is working more versus spending time with family, which accounts for 10.9% of respondents. The next most common pairs are working more versus leisure, at 10.0%, and taking a higher-paying job versus spending time with family, at 7.54%.

4.7 Fielding the Follow-Up Survey

Respondents were invited to participate in the follow-up survey approximately one month after the baseline survey. About 82.3% of subjects in the main analysis sample at baseline responded to the follow-up survey, and response rates were similar between treatment and

[30] This average, 2.47, corresponds to the choice framing; the corresponding value is slightly smaller, 2.31, under the satisfaction framing.

control groups.[31] The follow-up survey was substantially shorter than the baseline survey, with a median duration of about 6.20 minutes.[32] 94.9% of respondents rated it as very easy or somewhat easy to complete. The follow-up survey also included the video-code check, and the vast majority of subjects passed it.[33]

5 Empirical Results

5.1 Perceived MSI

In this section, we provide some descriptive evidence about self-MSI and average-MSI beliefs. Panel A of Figure 4 overlays the histograms of prior beliefs about the average-MSI (solid bars) and the self-MSI (hollow bars). We start with prior average-MSI, which has a mean of 5.20. That is, on average, respondents believe that a 20% income increase would raise the life satisfaction of an average person by 5.20 points on the 0–100 scale. There is substantial dispersion in beliefs across subjects: while mean prior average-MSI is 5.20, the 25th percentile is 3 and the 75th percentile is 7. The dispersion is broadly similar for self-MSI beliefs, but the key difference between the two is that the distribution of self-MSI beliefs is shifted to the right relative to the distribution of average-MSI beliefs: self-MSI beliefs are on average 1.60 points higher than average-MSI beliefs (6.80 vs. 5.20). The binned scatterplot from Panel B of Figure 4 shows that prior average-MSI and self-MSI beliefs are highly correlated with each other, but far from perfectly so: their correlation coefficient is 0.636 ($p<0.001$).[34]

One common concern with survey data that applies to these measures of prior beliefs is the role of measurement error. The last four panels of Figure 4 provide a sense of whether these belief measures contain meaningful signal rather than just noise, where Panels C and D correspond to average-MSI beliefs while Panels E and F correspond to self-MSI beliefs. All four panels are based on control-arm respondents who completed the follow-up survey, and the results concern the same set of respondents throughout, so that the within-baseline reliabilities (Panels C and E) and the one-month reliabilities (Panels D and F) are computed on identical samples and are directly comparable. In Panel C, for example, subjects were asked

[31] 82.1% in the control arm versus 82.6% in the treatment arm, with the difference being statistically indistinguishable from zero ($p=0.728$). The pre-treatment characteristics are also balanced across treatment arms within the follow-up sample—see Appendix C.3. Follow-up subjects are a bit different from the baseline respondents: they are older, more active on Prolific, and have somewhat lower baseline self-MSI beliefs; see Appendix C.2 for more details.

[32] The follow-up survey had a participation reward of $1.50.

[33] The follow-up survey included the same video-code check used at baseline to detect AI-assisted responses, but not the end-of-survey attention checks. 96.4% of follow-up respondents passed this check.

[34] This high correlation is consistent with the learning framework from Section 2, according to which respondents use their available data to learn jointly about self-MSI and average-MSI.

the same question about average-MSI twice during the baseline survey, but did not receive any information between the two elicitations. For that reason, if subjects are certain about average-MSI, we would expect them to provide the same answer both times. At the other extreme, if the answers to these two questions were uncorrelated, that would indicate that they are pure noise. Under the classical-measurement-error model, this test-retest correlation (or reliability coefficient) can be interpreted as the share of observed variation attributable to signal rather than noise. Within the baseline survey, the test-retest correlations are high: 0.729 for average-MSI beliefs and 0.876 for self-MSI beliefs. Thus, this simple reliability calculation suggests that about 72.9% of the variation in average-MSI beliefs and 87.6% of the variation in self-MSI beliefs reflects a meaningful signal. These reliabilities are comparable to, or higher than, standard benchmarks from survey measurement. For example, Krueger and Schkade (2008) report test-retest correlations of about 0.59 for life satisfaction and up to 0.90 for education and income.

At the same time, the within-baseline correlations may overstate reliability if some respondents remembered their earlier answers and repeated them for consistency. To address this concern, Panels D and F of Figure 4 report the correlation between baseline prior beliefs and the same belief measures elicited in the follow-up survey about one month later. This longer-horizon comparison has the opposite limitation: it mechanically understates reliability as a measure of signal, because respondents may genuinely revise their beliefs between the baseline and follow-up surveys. That is, over the course of a month, respondents may have encountered new information about the relationship between income and life satisfaction and updated their beliefs accordingly.[35] With this caveat in mind, the one-month test-retest correlations remain meaningfully positive: 0.345 for average-MSI beliefs and 0.457 for self-MSI beliefs. Thus, even using this more demanding comparison, the beliefs still contain substantial signal.

For additional evidence on reliability, we use the question from the follow-up survey about how confident respondents were in their average-MSI and self-MSI beliefs. According to the conceptual framework, individuals are forming these beliefs with imperfect data, so we would not expect them to be extremely confident. However, if everyone answered "not at all confident", we would worry that their answers are wild guesses and thus pure noise. The distribution of responses leans heavily towards high confidence: in the control group 24.9% reported being extremely confident in their self-MSI belief, 45.0% very confident, 24.3% moderately confident, 4.97% slightly confident, and only 0.78% not at all confident. Confidence is lower for average-MSI beliefs—see Appendix D.1 for more details. One interpretation of this

[35] In addition, for self-MSI beliefs, the fundamentals may themselves change: for example, a respondent who receives a large hospital bill or experiences another financial shock may have a genuinely different MSI one month later.

finding could be that individuals have more precise information about their own satisfaction and income than about other individuals' satisfaction and income.

Finally, we assess the reliability of prior beliefs by examining whether prior self-MSI predicts how individuals trade off income against other considerations. Such predictive power would suggest that these beliefs contain signal, even if they are noisy. The results are presented in Figure 5. Each of the four panels shows the relationship between prior self-MSI and a different type of choice. All four associations are statistically and economically significant. For example, Panel A shows that, on average, individuals with higher self-MSI beliefs are more willing to work longer hours in the work-hours job-choice scenario. More precisely, a 1-point increase in prior self-MSI is associated with a 0.027-point increase in willingness to work longer hours on the 1–4 scale ($p < 0.001$). Panels B and C show quantitatively and qualitatively consistent results for the income-versus-commute and income-versus-sleep choices, and Panel D shows a similar pattern for the real-world decision. These associations suggest that prior self-MSI beliefs are informative, but they do not establish causality. We examine the causal effect of self-MSI beliefs on choices in Section 5.5.

5.2 Biases in Perceived MSI

The first prediction from the conceptual framework is that individuals hold upward-biased self-MSI and average-MSI beliefs, with the bias weakly larger for self-MSI beliefs. Before proceeding, we must emphasize an important caveat that applies throughout the paper: our interpretation of differences between beliefs and the scientific benchmark as "belief biases" relies on the assumption that the scientific evidence provides an unbiased estimate of the true average-MSI. For example, one potential source of bias is the reporting scale: if respondents change how they use satisfaction scales as their circumstances change, changes in reported scores need not fully reflect changes in underlying satisfaction (Kaiser and Lepinteur, 2025; Prati and Senik, 2026). The benchmark we use as a signal is an estimate, and any error in it would affect the magnitude of the belief biases we report. For example, if the true average-MSI were +2 rather than +1, our estimate of mean overestimation in average-MSI beliefs would be one point smaller. The estimated treatment effects on reported beliefs and choices would remain unchanged. However, interpreting these responses as correcting misperceptions and improving decision-making depends on the accuracy of the benchmark.

The key evidence is shown in Panel A of Figure 4. The overlapping histograms show the distributions of prior average-MSI and self-MSI beliefs, respectively. In turn, the vertical line corresponds to the signal about average-MSI provided in the treatment message, which has a value of +1. We start with average-MSI beliefs because they can be compared more directly with the scientific benchmark. The vast majority (93.8%) of subjects' priors are

above the benchmark of +1 from the scientific evidence. If we compare mean prior average-MSI (5.20) to the benchmark (+1), this implies a mean overestimation of 4.20 points.[36] This mean overestimation is not only statistically significant ($p<0.001$) but also economically meaningful: the mean prior belief is 5.20 times the scientific benchmark.

To help interpret the magnitude of the bias in the average-MSI beliefs, recall that the hypothetical average person starts with life satisfaction of 75 points. An average-MSI belief of 15—the upper end of our response scale—therefore implies that a 20% income increase raises satisfaction by 20%, corresponding to an approximate income elasticity of life satisfaction of 1. In other words, an average-MSI belief of 15 implies that satisfaction rises approximately proportionally with income. By comparison, mean prior average-MSI of 5.20 implies an elasticity of approximately 0.35 ($\approx 6.9\%/20\%$), while the scientific benchmark of 1 point implies an elasticity of approximately 0.067 ($\approx 1.3\%/20\%$). Although respondents systematically overestimate average-MSI, they perceive satisfaction as inelastic with respect to income—that is, they expect satisfaction to increase less than proportionally with income.

To assess the bias in self-MSI beliefs, ideally we would observe the *true* MSI of each respondent and compare it to the respondent's prior self-MSI belief. Of course, it would be impossible to estimate the true MSI of each respondent with any reasonable degree of certainty. However, while we cannot estimate the individual bias of each respondent, with a simple assumption, we can easily estimate the *mean* bias in self-MSI beliefs across respondents. To be precise, borrowing the notation from Section 2, let $\tilde{\beta}_i$ denote respondent i's prior self-MSI belief and let β_i denote respondent i's true self-MSI. If we average the individual biases, $\tilde{\beta}_i - \beta_i$, across respondents, we obtain the following:

$$\frac{1}{N}\sum_i \left(\tilde{\beta}_i - \beta_i\right) = \frac{1}{N}\sum_i \tilde{\beta}_i - \frac{1}{N}\sum_i \beta_i \tag{9}$$

The key assumption we need is that the +1 benchmark from the scientific evidence is an unbiased estimate of the mean MSI in the subject pool. In other words, we need to assume that mean MSI across subjects is comparable to the mean MSI from the scientific evidence. Indeed, this is an assumption we can test with income and life satisfaction data we collected in our survey, and it seems to be broadly consistent with the data—for more details, see Appendix C.4.

In sum, while we cannot assess biases at the individual level, we can assess the mean bias in self-MSI beliefs by just comparing mean prior self-MSI to the +1 benchmark from the scientific evidence. Panel A of Figure 4 shows that mean prior self-MSI is 6.80, implying a mean bias of 5.80 points. As with average-MSI beliefs, the mean bias in self-MSI beliefs

[36] Misperceptions about average-MSI are consistent with a large body of evidence that individuals hold systematic misperceptions about others (Bursztyn and Yang, 2022).

is statistically significant and economically meaningful.[37] Moreover, the mean bias in self-MSI beliefs (5.80 points) is (weakly) larger than the mean bias in average-MSI beliefs (4.20 points), with the difference being statistically significant ($p<0.001$).[38]

Finally, we discuss two points that are relevant to interpreting these findings. First, in our simple framework, individuals do not learn about their own adaptation neglect. In practice, however, experience and education may help individuals learn about their biases (Ali, 2011).[39] We find that average-MSI and self-MSI beliefs are lower among older and more educated respondents, which could be consistent with such learning—see Appendix D.2 for more details. Second, one might ask whether respondents perceive the hypothetical income increase as temporary or permanent. If consumption is a key mechanism through which income raises satisfaction, this distinction may matter. To the extent that individuals save part of a temporary income increase for consumption in future years, some of the associated satisfaction gains may also occur only later. Our survey therefore aimed at eliciting beliefs about a permanent increase in income.[40] If respondents instead interpreted the increase as temporary and expected smaller satisfaction gains from temporary than from permanent increases, they would have reported lower MSI beliefs than under the intended interpretation. In that case, respondents' true overestimation of the marginal satisfaction from income would be even greater than we document here.

5.3 Belief Updating and Persistence

The second prediction of the conceptual framework is that the provision of accurate information lowers, on average, both average-MSI and self-MSI beliefs, with a larger reduction in average-MSI beliefs. The key evidence is presented in Figure 6, which shows the distribution of posterior beliefs by treatment status, separately for average-MSI and self-MSI beliefs.[41] Panel A shows posterior average-MSI beliefs. Mean posterior average-MSI is 5.81 in the

[37] Self-MSI beliefs may be subject to social desirability bias: respondents may underreport their beliefs to avoid appearing overly materialistic. Survey anonymity may help mitigate this concern. In any case, any remaining downward bias would mean that both the mean overestimation of self-MSI and the gap between self-MSI and average-MSI beliefs are even larger than we document here.

[38] As explained in Section 4, the main analysis sample excludes respondents with an extreme prior average-MSI belief (-1 or 15) but not those with an extreme prior self-MSI belief. Thus, this exclusion criterion can mechanically contribute to the gap between self-MSI and average-MSI beliefs. However, even without this exclusion, the bias in self-MSI beliefs remains larger than the bias in average-MSI beliefs, with a difference of 1.46 points ($p<0.001$).

[39] Errors can nevertheless persist even when agents learn over time (Gagnon-Bartsch et al., 2026).

[40] Neither question explicitly specified whether the income increase would persist beyond that year. However, the average-MSI question described income "per year", which was intended to encourage respondents to think in terms of permanent income.

[41] As a falsification test, Figure D.4 reproduces Panels A and B of Figure 6, but uses prior beliefs instead of posterior beliefs. As expected, we find no treatment effects on prior beliefs.

control arm and 2.50 in the treatment arm, implying a treatment effect of -3.32 points, or about 0.967 SD of the control-group posterior distribution. Panel B shows the analogous comparison for posterior self-MSI beliefs: the mean is 6.95 in the control arm and 4.23 in the treatment arm, implying a treatment effect of -2.72 points, or about 0.623 SD of the control-group posterior distribution. In other words, providing information about average-MSI reduced both average-MSI and self-MSI beliefs on average, with a larger reduction in average-MSI beliefs. The sizable belief updating is consistent with high engagement with the information treatment. All treated subjects played the video, as required to advance to the next screen; 87.6% watched it essentially to the end, and the median treated respondent spent 233 seconds on the information-treatment screens.[42]

Panels C and D of Figure 6 show the corresponding follow-up beliefs elicited about one month later. This follow-up comparison provides a stronger test of belief change: the immediate posterior elicitation took place just after the information-treatment stage, when the signal was highly salient, while the follow-up beliefs capture the component of updating that respondents retain after leaving the survey. The effects remain sizable. About one-third of the initial treatment effect persists one month later (34.8% of the baseline effect on average-MSI beliefs and 36.0% of the effect on self-MSI beliefs), and treated subjects continue to report lower average-MSI and self-MSI beliefs.[43] Thus, the follow-up evidence suggests that the information treatment produced persistent belief updating, rather than only short-run responses to information that has just been presented. The magnitude is also in line with prior information-provision experiments. For example, Cavallo et al. (2017) find that roughly 45–48% of belief updating about inflation expectations persisted in follow-up surveys conducted two to four months later.

Appendix D.3 examines heterogeneity in belief updating by respondents' prior beliefs. The baseline survey suggests some heterogeneity: respondents with higher prior beliefs appear to update more strongly in the direction of the information. However, the follow-up survey suggests that most of this heterogeneity was spurious or transitory, as the persistent treatment effects one month later are more similar across the prior-belief distribution.

[42]Moreover, among treated respondents in the main analysis sample, 21.1% clicked "Show more" on treatment page 1 and 10.6% clicked "Show more" on treatment page 2. Overall, 26.7% clicked "Show more" on at least one of the two information-module "Show more" links. A negligible share clicked on the external links: 0.66% clicked the *Review of Economic Studies* link and 0.44% clicked the *New York Times* link.

[43]The follow-up survey included questions about respondents' confidence in their average-MSI and self-MSI beliefs. We find no evidence that the treatment significantly increased reported confidence in either belief—see Appendix D.1 for details.

5.4 Treatment Effects on Choices

The third prediction of the conceptual framework is that the provision of accurate information lowers, on average, the willingness to choose the income-generating action. The key evidence is presented in Figure 7, which compares the mean belief and choice outcomes between treatment and control groups. Each panel plots the mean outcome for the control and treatment arms. While the goal of this figure is to summarize the effects on choices, Panel A begins with posterior beliefs as a benchmark, and Panels B through F show the effects on the different behavioral outcomes. The results from Panel A of Figure 7 have been discussed already in Section 5.3 above. Most importantly, the information treatment reduced the mean posterior self-MSI by 2.72 satisfaction points ($p<0.001$).[44]

Panel B of Figure 7 corresponds to the decision involving 20% higher earnings at the expense of 20% more work hours. When asked what they would choose, subjects in the treatment arm are less inclined to choose the higher-income option: the mean falls from 2.45 in the control arm to 2.36 in the treatment arm, on the 1–4 scale coded so that 1 indicates definitely choosing the lower-income option and 4 indicates definitely choosing the higher-income option. This corresponds to an estimated effect of -0.097 ($p=0.020$). Panel B also reports the corresponding results but using the outcome under the satisfaction framing instead of the choice framing. In the satisfaction-framing version, the decline is very similar, from 2.36 to 2.27, with an estimated effect of -0.090 ($p=0.027$). The similarity of the results across the choice and satisfaction framings is not mechanical: responses under the two framings are highly correlated but not perfectly so, with correlations ranging from 0.695 to 0.784—see Appendix D.6 for details.[45]

Panel C of Figure 7 considers the analogous trade-off involving commuting time. Here, again, the treatment moves responses away from the higher-income option. In the choice-framing version, the mean falls from 1.92 to 1.85, an effect of -0.071 that is statistically significant ($p=0.064$). In the satisfaction-framing version, the mean falls from 1.88 to 1.81, also an effect of -0.071 that is statistically significant ($p=0.056$). Panel D of Figure 7 shows the trade-off involving sleep. The point estimates again go in the same negative direction, but they are smaller and less precisely estimated. In the choice-framing version, the treatment reduces the outcome from 2.81 to 2.75, an effect of the same order of magnitude (-0.066) but borderline statistically insignificant ($p=0.130$). In the satisfaction-framing version, the mean

[44] For respondents' decision-making, the relevant belief is the posterior self-MSI. However, for the sake of transparency, Panel A also shows that the information treatment also reduced the mean posterior average-MSI by 3.32 points ($p<0.001$).

[45] These results are consistent with the strong but imperfect correspondence between stated choices and predicted life satisfaction documented by Benjamin et al. (2012, p. 2103): in their Denver sample, a univariate regression using responses demeaned within scenarios yields an $R^2 = 0.65$, implying a correlation of approximately 0.81.

falls from 2.65 to 2.61, an effect that is still negative but smaller (-0.033) and statistically insignificant (p=0.447). To maximize statistical power, Panel E of Figure 7 averages the three outcomes from Panels B through D. The treatment reduces the choice-framing index from 2.39 to 2.32, a statistically significant (p=0.007) effect of -0.078. For the satisfaction-framing index, the mean falls from 2.30 to 2.23, a statistically significant effect of -0.065 (p=0.027).

The consistency across the three job-choice scenarios is informative because these scenarios capture distinct trade-offs involving income. For example, pairwise correlations between the three job-choice outcomes range from 0.196 to 0.313 under the choice framing—see Appendix D.6 for details. In other words, if an individual is willing to trade off higher income against one thing (e.g., a longer commute), that does not always mean that the individual would be willing to trade off higher income against another thing (e.g., less sleep). Thus, similar effects across these outcomes are not a mechanical consequence of respondents making nearly identical choices.

Panel F of Figure 7 corresponds to the real-world decision. The estimated treatment effects are still negative and of the same order of magnitude as for the other outcomes in Panels B–E, but less precisely estimated. For the choice-framing version of the real-world decision, the mean falls from 2.47 to 2.41, an effect of -0.066 that is not statistically significant (p=0.171). For the satisfaction-framing version, the mean falls from 2.31 to 2.21, a statistically significant effect of -0.096 (p=0.047). The signs and sizes are consistent with the job-choice outcomes, though the estimates are noisier.

Taken together, the pattern across Panels B through E of Figure 7 is reassuring. Even when some outcome-specific estimates are not statistically significant, the effects always point in the same direction and are of the same order of magnitude. Overall, the evidence suggests that the treatment shifted respondents away from higher-income options. The real-world decision outcome is also far from perfectly correlated with the three job-choice outcomes—see Appendix D.6. Thus, the similarity between the results for real-world decisions and job-choice scenarios is also informative and not just mechanical.

There is one additional outcome—the number of surveys completed on Prolific—for which data limitations make the results more speculative. The full set of results is presented in Appendix D.10 and summarized below. Our original intention was to test whether the treatment affected the number of surveys participants subsequently completed. In the follow-up survey, we asked participants to report how many surveys they had completed during the previous 30 days, a period that roughly covers the time since the baseline survey. Our hypothesis was that treated participants might complete fewer surveys because they placed less value on additional income. However, we found that participants generally had poor recall of how many surveys they had completed in the past. In the follow-up survey, we

addressed this problem by asking participants to retrieve the exact number from their Prolific dashboards, but we were able to obtain this measure for only a subset of the sample. With this caveat, we nevertheless find suggestive evidence of treatment effects on this additional outcome: during the month following the baseline survey, there is a negative treatment effect on the number of Prolific surveys completed.

5.5 Effects of Self-MSI Beliefs on Choices

The results from the previous section capture the intention-to-treat effects of providing information. In this section, we focus instead on measuring the causal effect of beliefs on choices. We start with a simple OLS regression of the choice outcome (C_i^k) on posterior self-MSI beliefs ($MSI_{i,post}^{self}$):

$$C_i^k = \mu_0^k + \mu_{MSI}^k \cdot MSI_{i,post}^{self} + u_i^k \tag{10}$$

The key coefficient of interest is μ_{MSI}^k, which captures the behavioral effect of increasing posterior self-MSI by one point. As a starting point, this simple OLS regression leverages both the experimental and non-experimental variation in posterior self-MSI beliefs. Although this approach maximizes statistical power, the use of non-experimental variation may introduce omitted-variable bias.

The results from the OLS regressions are presented in Panel A of Table 2. The coefficients are reported separately for the choice-framing and the satisfaction-framing outcomes, and we start by discussing the former. Column (1) of Panel A implies that a 1-point decrease in posterior self-MSI is associated with a 0.025-point decrease in willingness to choose the higher-income option with longer work hours (on the 1–4 scale), corresponding to about 0.023 SD of this outcome. Columns (2)–(3) show that the results are qualitatively and quantitatively similar for the other two job-choice scenarios, and also when we use the index comprising all three scenarios (column (4)). Column (5) also shows similar results when using the real-world decision outcome. To put these magnitudes in further perspective, a one-standard-deviation increase in posterior self-MSI—4.37 points in the control group—is associated with an increase of approximately 0.15 SD in the job-choice index. The corresponding associations range from 0.09 to 0.13 SD across the three individual job-choice outcomes and equal approximately 0.12 SD for the real-world decision outcome. Thus, the relationship between self-MSI beliefs and choices is also economically meaningful. Moreover, the second row of Panel A of Table 2 shows the corresponding results for the satisfaction framing, which are very similar to the choice-framing results, both in magnitude and in statistical significance.

Panel B of Table 2 estimates the effect of posterior self-MSI on behavior using a 2SLS

specification, with the treatment indicator (T_i) as the excluded instrument. This specification isolates the variation in posterior self-MSI beliefs induced by the randomized information provision. Because treatment was randomly assigned, this induced variation in posterior self-MSI is exogenous. The causal interpretation still requires an exclusion restriction: the information treatment must affect choices only through changes in posterior beliefs ($MSI^{self}_{i,post}$).[46]

Starting with the choice-framing outcomes, column (1) of Table 2 shows that a 1-point decrease in posterior self-MSI causes a 0.035-point decrease in willingness to choose higher income over fewer work hours on the 1–4 scale. Relative to the control-group SD, this corresponds to about 0.033 SD. Column (2) shows a similar effect for commuting: a 1-point decrease in posterior self-MSI causes a 0.026-point decrease in willingness to choose higher income over a shorter commute, or about 0.025 SD. Column (3) shows an effect of 0.024 points for sleep, equal to about 0.021 SD. Column (4), which averages the three job-choice outcomes, implies an effect of 0.028 points, or about 0.037 SD. Finally, column (5) shows a similar effect for the real-world decision outcome, with a 0.024-point decline, or about 0.024 SD.

Under the satisfaction framing, the 2SLS estimates are again positive, implying that reductions in posterior self-MSI shift respondents' satisfaction comparisons away from the higher-income option. The results under the satisfaction framing are also similar in magnitude to the choice-framing estimates. The corresponding 2SLS coefficients are 0.033 for work hours, 0.026 for commute, 0.012 for sleep, 0.024 for the index, and 0.035 for the real-world decision outcome. In SD units, these correspond to approximately 0.031, 0.026, 0.011, 0.031, and 0.035 SD, respectively. The significance pattern is also similar under the satisfaction and choice framings, with the exception of the real-world decision, which is statistically significant under the satisfaction framing but not under the choice framing.

The comparison between Panels A and B of Table 2 is useful. The key trade-off is that Panel B has a cleaner causal interpretation, because it isolates the component of posterior self-MSI shifted by the randomized information treatment, but it is also less precisely estimated because it uses only a small share of the overall variation in posterior self-MSI: the shocks induced by the experiment. Despite this loss of precision, the estimates in Panels A and B are surprisingly similar in sign and size, which is reassuring. While some specific coefficients are statistically insignificant, the same result appears across all outcomes, framings, and baseline specifications: lower posterior self-MSI beliefs reduce respondents' willingness to

[46] As in standard 2SLS settings, when treatment effects vary across individuals, the estimate captures a local average treatment effect of beliefs, in the sense of Imbens and Angrist (1994). In this context, that means the estimated treatment effect is weighted more heavily toward individuals whose beliefs changed the most in response to the information. This interpretation should be treated with caution, however, because it requires monotonicity, which may not hold in practice given that some respondents' prior average-MSI beliefs lie below the signal.

choose the higher-income option. Appendix D.5 confirms that the job-choice index estimates are robust to a battery of alternative samples and specifications. Regarding the real-world decisions, however, the results depend more strongly on some of the criteria used for coding this outcome, for example, excluding observations with ambiguous trade-offs—for more details, see Appendix D.8.

Next, we evaluate whether the experimental effects persisted at follow-up. The relevant results are presented in Panel C of Table 2. This panel is analogous to Panel B, but both the outcome and the endogenous variable are measured at follow-up. Specifically, the dependent variable is willingness to choose the higher-income option at follow-up, the endogenous variable is posterior self-MSI measured at follow-up, and the instrument is again the randomized treatment indicator. Thus, Panel C asks whether the experimentally induced changes in self-MSI beliefs continue affecting choices at follow-up.

Two features of the follow-up data reduce precision. First, attrition lowers sample sizes. For the job-choice scenarios, the sample size declines by 18.2 percent, from 2,721 respondents at baseline to 2,225 at follow-up. For the real-world decision outcome, it falls by 19.1 percent, from 1,711 respondents at baseline to 1,384 at follow-up. Second, the first stage is weaker at follow-up. Although the treatment effect on self-MSI beliefs remains statistically significant, it is only partially persistent: the effect falls from 2.72 points at baseline to about 0.980 points at follow-up, roughly one-third of the baseline effect. Together, the smaller sample and weaker first stage make the follow-up 2SLS estimates substantially less precise. With those caveats in mind, the point estimates in Panel C of Table 2 remain broadly consistent with those in Panel B, although two of the five (commute and the real-world decision) are slightly negative. In column (1) of the table, the coefficient is 0.026 in Panel C, compared with 0.035 in Panel B. These magnitudes are quite similar. However, the standard error rises sharply: from 0.015 in Panel B to 0.046 in Panel C, roughly three times as large. This means that, at follow-up, we do not have enough statistical power to detect effects of the size estimated in Panel B. Taken together across all outcomes, the follow-up estimates are broadly consistent with the baseline causal estimates, but too noisy to draw strong conclusions about persistence. The results are therefore best read as suggestive: they do not contradict the baseline pattern, but they also do not allow us to estimate the follow-up effects with enough precision.

Lastly, a common concern with 2SLS estimates is weak-instrument bias (Stock et al., 2002). Given the strong effect of the treatment on posterior self-MSI, this is unlikely to be a concern in our application. For a formal assessment, the bottom of Table 2 reports the Kleibergen-Paap *rk* Wald F-statistics for each 2SLS specification, a standard measure

of instrument strength.[47] Following the guideline of Staiger and Stock (1997), F-statistics of 10 or higher indicate that weak identification is not a serious concern. Our reported values are well above this threshold, confirming that weak instruments are not a concern. Appendix D.4 shows that this first-stage effect of treatment on posterior self-MSI is large and highly significant across all alternative samples and specifications we consider.

6 Conclusions

We conducted a pre-registered information-provision experiment to study whether individuals misperceive MSI and whether these misperceptions affect choices. The experiment elicits self-MSI and average-MSI beliefs and randomly assigns respondents to receive scientific evidence about average-MSI. We find that respondents' average-MSI and self-MSI beliefs lie above the scientific benchmark, with a larger gap for self-MSI beliefs. When presented with scientific evidence showing that the effect of income on life satisfaction is smaller than they expected, respondents revise their beliefs downward. A follow-up survey shows that these belief changes are not simply transitory, as one-third of the treatment effects persist one month later. Most importantly, the belief changes translate into choices: treated respondents become less willing to choose the higher-income option when presented with hypothetical and real-world trade-offs between income and another dimension of life they value. Taken together, the results suggest that beliefs about the income-satisfaction relationship can be consequential for economic decision-making.

References

Affonso, F. M. (2026). Brief Commentary: A Framework for Detecting AI Agents in Online Research. *Journal of Consumer Research*, ucag006.

Aknin, L. B., M. I. Norton, and E. W. Dunn (2009). From Wealth to Well-Being? Money Matters, but Less than People Think. *The Journal of Positive Psychology 4*(6), 523–527.

Ali, S. N. (2011). Learning Self-Control. *The Quarterly Journal of Economics 126*(2), 857–893.

Allcott, H., L. Braghieri, S. Eichmeyer, and M. Gentzkow (2020). The Welfare Effects of Social Media. *American Economic Review 110*(3), 629–676.

Apouey, B. and A. E. Clark (2015). Winning Big but Feeling No Better? The Effect of Lottery Prizes on Physical and Mental Health. *Health Economics 24*(5), 516–538.

Baum, C. F., M. E. Schaffer, and S. Stillman (2007). Enhanced Routines for Instrumental Variables/GMM Estimation and Testing. *The Stata Journal 7*(4), 465–506.

Becker, G. S. and L. Rayo (2008). Comment on "Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox" by Betsey Stevenson and Justin Wolfers. *Brookings Papers on Economic Activity 39*(1), 88–95. Spring.

[47] Although the conventional rule of thumb is based on the Cragg-Donald statistic, which assumes homoskedastic errors, Baum et al. (2007) recommend using the Kleibergen-Paap statistic as its robust counterpart.

Benjamin, D. J., O. Heffetz, M. S. Kimball, and A. Rees-Jones (2012). What Do You Think Would Make You Happier? What Do You Think You Would Choose? *American Economic Review 102*(5), 2083–2110.

Benjamin, D. J., O. Heffetz, M. S. Kimball, and A. Rees-Jones (2014). Can Marginal Rates of Substitution Be Inferred from Happiness Data? Evidence from Residency Choices. *American Economic Review 104*(11), 3498–3528.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Information Science and Statistics. New York, NY: Springer.

Bursztyn, L. and D. Y. Yang (2022). Misperceptions about Others. *Annual Review of Economics 14*(1), 425–452.

Cavallo, A., G. Cruces, and R. Perez-Truglia (2017). Inflation Expectations, Learning, and Supermarket Prices: Evidence from Survey Experiments. *American Economic Journal: Macroeconomics 9*(3), 1–35.

Çelebi, C., C. Exley, S. Harrs, H. Kivimaki, M. Serra-Garcia, and J. Yusof (2026). Mission Possible: The Collection of High-Quality Online Data. *CESifo Working Paper No. 12535*.

Chopra, F. and I. Haaland (2023). Conducting Qualitative Interviews with AI. CESifo Working Paper 10666, CESifo.

Clark, A. E., S. Fl'eche, R. Layard, N. Powdthavee, and G. Ward (2018). *The Origins of Happiness: The Science of Well-Being over the Life-Course*. Princeton, NJ: Princeton University Press.

Clark, A. E., P. Frijters, and M. A. Shields (2008). Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles. *Journal of Economic Literature 46*(1), 95–144.

Clark, A. E. and Y. Georgellis (2013). Back to Baseline in Britain: Adaptation in the British Household Panel Survey. *Economica 80*(319), 496–512.

Clark, A. E. and C. Senik (2010). Who Compares to Whom? The Anatomy of Income Comparisons in Europe. *The Economic Journal 120*(544), 573–594.

Cone, J. and T. Gilovich (2010). Understanding Money's Limits: People's Beliefs about the Income-Happiness Correlation. *The Journal of Positive Psychology 5*(4), 294–301.

Costello, T. H., G. Pennycook, and D. G. Rand (2024). Durably Reducing Conspiracy Beliefs through Dialogues with AI. *Science 385*(6714), eadq1814.

Cullen, Z. and R. Perez-Truglia (2022). How Much Does Your Boss Make? The Effects of Salary Comparisons. *Journal of Political Economy 130*(3), 766–822.

Di Tella, R., J. Haisken-De New, and R. MacCulloch (2010). Happiness Adaptation to Income and to Status in an Individual Panel. *Journal of Economic Behavior & Organization 76*(3), 834–852.

Dwyer, R. J. and E. W. Dunn (2022). Wealth Redistribution Promotes Happiness. *Proceedings of the National Academy of Sciences 119*(46), e2211123119.

Epstein, L. G. and M. Schneider (2007). Learning under Ambiguity. *The Review of Economic Studies 74*(4), 1275–1303.

Frey, B. S. and A. Stutzer (2014). Economic Consequences of Mispredicting Utility. *Journal of Happiness Studies 15*(4), 937–956.

Fuster, A., R. Perez-Truglia, M. Wiederholt, and B. Zafar (2022). Expectations with Endogenous Information Acquisition: An Experimental Investigation. *Review of Economics and Statistics 104*(5), 1059–1078.

Gagnon-Bartsch, T., M. Rabin, and J. Schwartzstein (2026). Channeled Attention and Stable Errors. *The Quarterly Journal of Economics*. Forthcoming.

Galiani, S., P. J. Gertler, and R. Undurraga (2018). The Half-Life of Happiness: Hedonic Adaptation in the Subjective Well-Being of Poor Slum Dwellers to the Satisfaction of Basic Housing Needs. *Journal of the European Economic Association 16*(4), 1189–1233.

Geiecke, F. and X. Jaravel (2026). Conversations at Scale: Robust AI-Led Interviews with a Simple

Open-Source Platform. *The Review of Economic Studies*. Forthcoming.

Giaccobasso, M., B. Nathan, R. Perez-Truglia, and A. Zentner (2025). Where Do My Tax Dollars Go? Tax Morale Effects of Perceived Government Spending. *American Economic Journal: Applied Economics 17*(4), 223–259.

Gordon, A. (2026). Authenticity Checks. https://www.prolific.com/resources/authenticity-checks-how-we-tested-the-most-accurate-method-for-identifying-agentic-ai, accessed May 18, 2026.

Grootendorst, M. (2022). BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure. *arXiv Preprint arXiv:2203.05794*.

Haushofer, J. and J. 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.

Hsee, C. K., Y. Yang, N. Li, and L. Shen (2009). Wealth, Warmth, and Well-Being: Whether Happiness Is Relative or Absolute Depends on Whether It Is about Money, Acquisition, or Consumption. *Journal of Marketing Research 46*(3), 396–409.

Imbens, G. W. and J. D. Angrist (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica 62*(2), 467–475.

Kahneman, D. and A. Deaton (2010). High Income Improves Evaluation of Life but Not Emotional Well-Being. *Proceedings of the National Academy of Sciences 107*(38), 16489–16493.

Kahneman, D. and R. H. Thaler (2006). Anomalies: Utility Maximization and Experienced Utility. *Journal of Economic Perspectives 20*(1), 221–234.

Kaiser, C. and A. Lepinteur (2025). Measuring the Unmeasurable? Systematic Evidence on Scale Transformations in Subjective Survey Data. *Wellbeing Research Centre Working Paper No. 2503*.

Killingsworth, M. A., D. Kahneman, and B. Mellers (2023). Income and Emotional Well-Being: A Conflict Resolved. *Proceedings of the National Academy of Sciences 120*(10), e2208661120.

Krueger, A. B. and D. A. Schkade (2008). The Reliability of Subjective Well-Being Measures. *Journal of Public Economics 92*(8–9), 1833–1845.

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies 76*(3), 1071–1102.

Lindqvist, E., R. Östling, and D. Cesarini (2020). Long-Run Effects of Lottery Wealth on Psychological Well-Being. *The Review of Economic Studies 87*(6), 2703–2726.

Loewenstein, G., T. O'Donoghue, and M. Rabin (2003). Projection Bias in Predicting Future Utility. *The Quarterly Journal of Economics 118*(4), 1209–1248.

Luttmer, E. F. P. (2005). Neighbors as Negatives: Relative Earnings and Well-Being. *The Quarterly Journal of Economics 120*(3), 963–1002.

Prati, A. and C. Senik (2026). Is It Possible to Raise National Happiness? *Journal of Public Economics 257*, 105619.

Rayo, L. and G. S. Becker (2007). Evolutionary Efficiency and Happiness. *Journal of Political Economy 115*(2), 302–337.

Schwarz, N. and J. Xu (2011). Why Don't We Learn from Poor Choices? The Consistency of Expectation, Choice, and Memory Clouds the Lessons of Experience. *Journal of Consumer Psychology 21*(2), 142–145.

Staiger, D. and J. H. Stock (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica 65*(3), 557–586.

Stevenson, B. and J. Wolfers (2013). Subjective Well-Being and Income: Is There Any Evidence of Satiation? *American Economic Review 103*(3), 598–604.

Stock, J. H., J. H. Wright, and M. Yogo (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics 20*(4), 518–529.

Stutzer, A. (2004). The Role of Income Aspirations in Individual Happiness. *Journal of Economic Behavior & Organization 54*(1), 89–109.

Vivalt, E., E. Rhodes, A. W. Bartik, D. E. Broockman, P. Krause, and S. Miller (2026). The Employment Effects of a Guaranteed Income: Experimental Evidence from Two U.S. States. *The Quarterly Journal of Economics 141*(4), 2675–2734.

Wilson, T. D. and D. T. Gilbert (2003). Affective Forecasting. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology*, Volume 35, pp. 345–411. Academic Press.

Figure 1: Treatment Screens

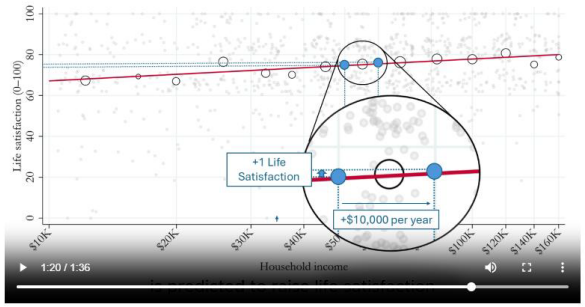
Panel A: First Treatment Screen

Income and Life Satisfaction

There is a large academic literature in economics and other social sciences examining the effects of income on subjective well-being, including happiness and life satisfaction. Across many studies conducted over several decades, a common conclusion is that income does affect happiness and life satisfaction, but that the effect is modest in size.

To illustrate the kind of relationship these studies typically find, we present evidence from the General Social Survey, a large U.S. survey that interviews thousands of Americans each year. In these data, a 20% increase in annual income—such as from $50,000 to $60,000—is associated with a 1-point increase in life satisfaction on a 0–100 scale in the same year.

Please take a moment to watch the video below for a more detailed explanation:



Panel B: Second Treatment Screen

What Science Says on the Causal Effects of Income on Life Satisfaction

In the previous page, we showed you a simple correlation using real survey data. However, more rigorous analyses that approximate scientific experimental studies find surprisingly similar numbers.

Long-Run Effects of Lottery Wealth on Psychological Well-Being

A landmark study by Economics Professors Lindqvist, Östling, and Cesarini study the effects of lottery winnings on life satisfaction. Their research study was peer-reviewed and published at *The Review of Economic Studies*, which is widely regarded as one of the best journals in the economics profession. It has also been featured in international media outlets such as The New York Times. Exploiting the random allocation of lottery prizes, they track winners for up to 22 years. They find that large windfalls lead to persistent increases in overall life satisfaction 5-22 years after the event, with no evidence of the effects vanishing over time. By converting the lottery prizes effects into an equivalent annual income increase lasting 20 years, a 20% raise in yearly income causes life satisfaction to increase by almost 1 point (on the 0–100 scale), which is similar to our estimates from the previous page.

Panel C: Third Treatment Screen

Why Higher Income Doesn't Always Bring More Satisfaction

Social scientists have conducted hundreds of studies to understand why the effect of income on life satisfaction is often modest. Many explanations have been proposed; below are some common explanations.

1 Hedonic Adaptation

People often get used to improvements.

One reason is that people often adapt to improvements in their circumstances. A raise, bonus, or windfall may feel significant at first, but over time a higher level of consumption becomes the new normal. In the well-being literature, this process is usually called hedonic adaptation. A new car, for example, may be exciting initially, but that satisfaction tends to fade, and after a few months one may feel much the same as while driving the old car.

Read more

2 Rising Aspirations

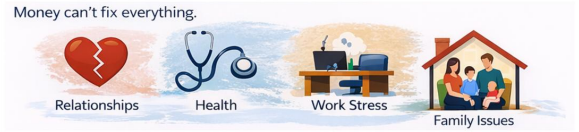
As we earn more, our sense of "enough" also increases.

A closely related explanation is that income aspirations rise with income itself. Survey-based research shows that as people earn more, they also tend to report a higher level of income as "sufficient." Intuitively, for someone earning $50,000 a year, $100,000 may seem more than sufficient; yet if that person eventually reaches $100,000, it may not be long before $200,000 comes to seem like what they need to live comfortably.

Read more

3 Limited Impact on Life Satisfaction

Money can't fix everything.



A third reason is that higher income mainly relieves certain problems, especially financial strain, rather than transforming every aspect of life. More income can reduce worry about bills, debt, and emergencies, which can in turn improve life satisfaction. But many other important determinants of life satisfaction — such as health, relationships, working conditions, loneliness, purpose, and family conflict — do not disappear simply because income rises. Extra income can therefore help, but only to a limited extent.

Read more

Notes: The information treatment consisted of three consecutive screens, with one screen shown in each panel. For full-page screenshots of each screen, see Appendix E.

Figure 2: Examples of Choice Elicitation Modules

Panel A: Job-Choice Module

Please consider the following two job opportunities. One offers higher pay but requires longer working hours. Aside from the differences summarized below, the two options are identical in all other respects.

If you were limited to these two options, which do you think you would choose?

	Offer A		Offer B	
Position Title	Lead Operations Processor		Lead Operations Processor	
Weekly Pay	$1,173		$1,408	
Weekly Hours Worked	40 hours		48 hours	
	○ definitely choose A	○ possibly choose A	○ possibly choose B	○ definitely choose B

Panel B: Real-World Decision Module

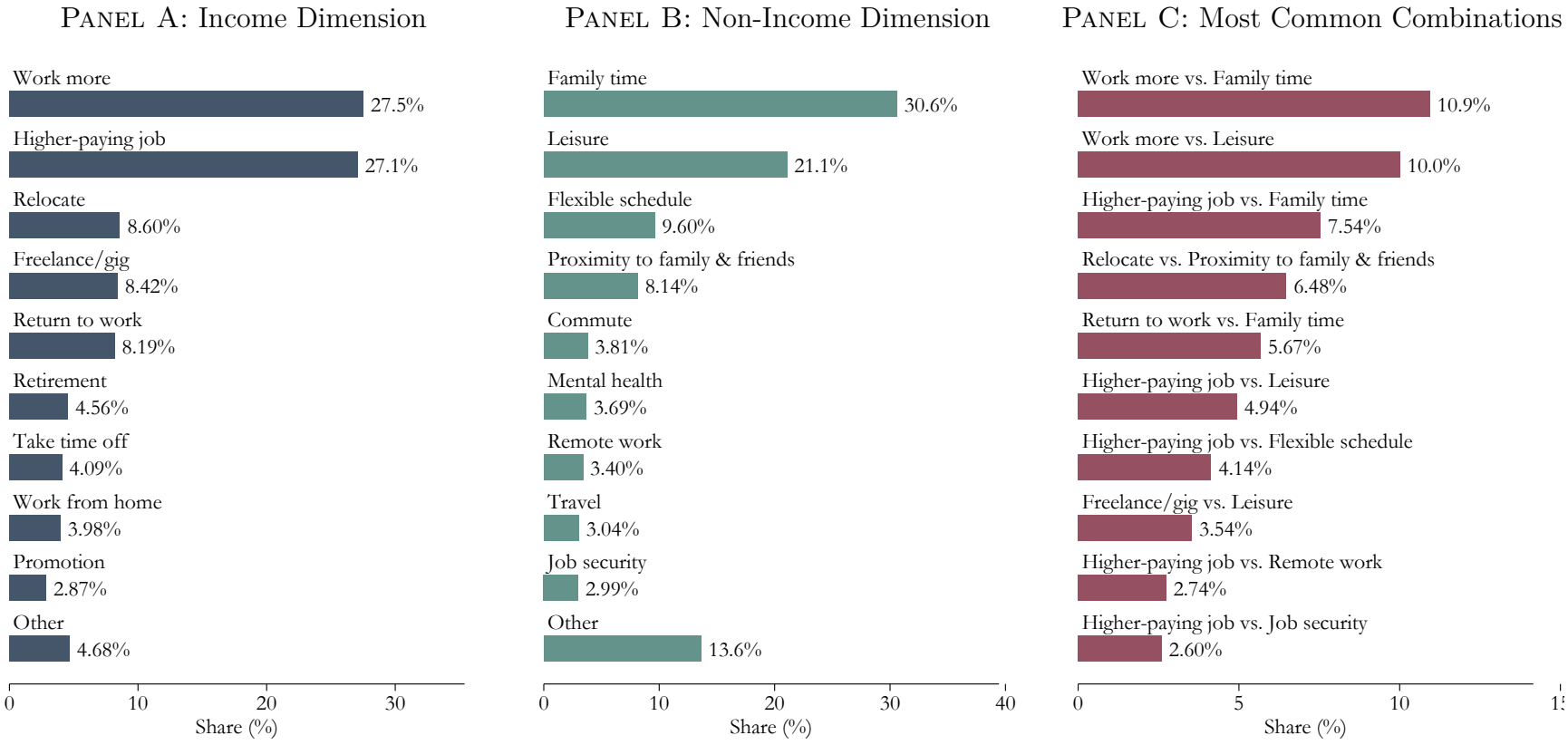
Let's start with the decision you discussed with the AI assistant earlier. If you work extra hours, you would earn more money for the wedding, but you would spend less time with your family and risk burnout.

If you were limited to these two options, which do you think you would choose?

Option A: Work extra hours for more wedding money		Option B: Do not work extra hours	
○ definitely choose A	○ possibly choose A	○ possibly choose B	○ definitely choose B

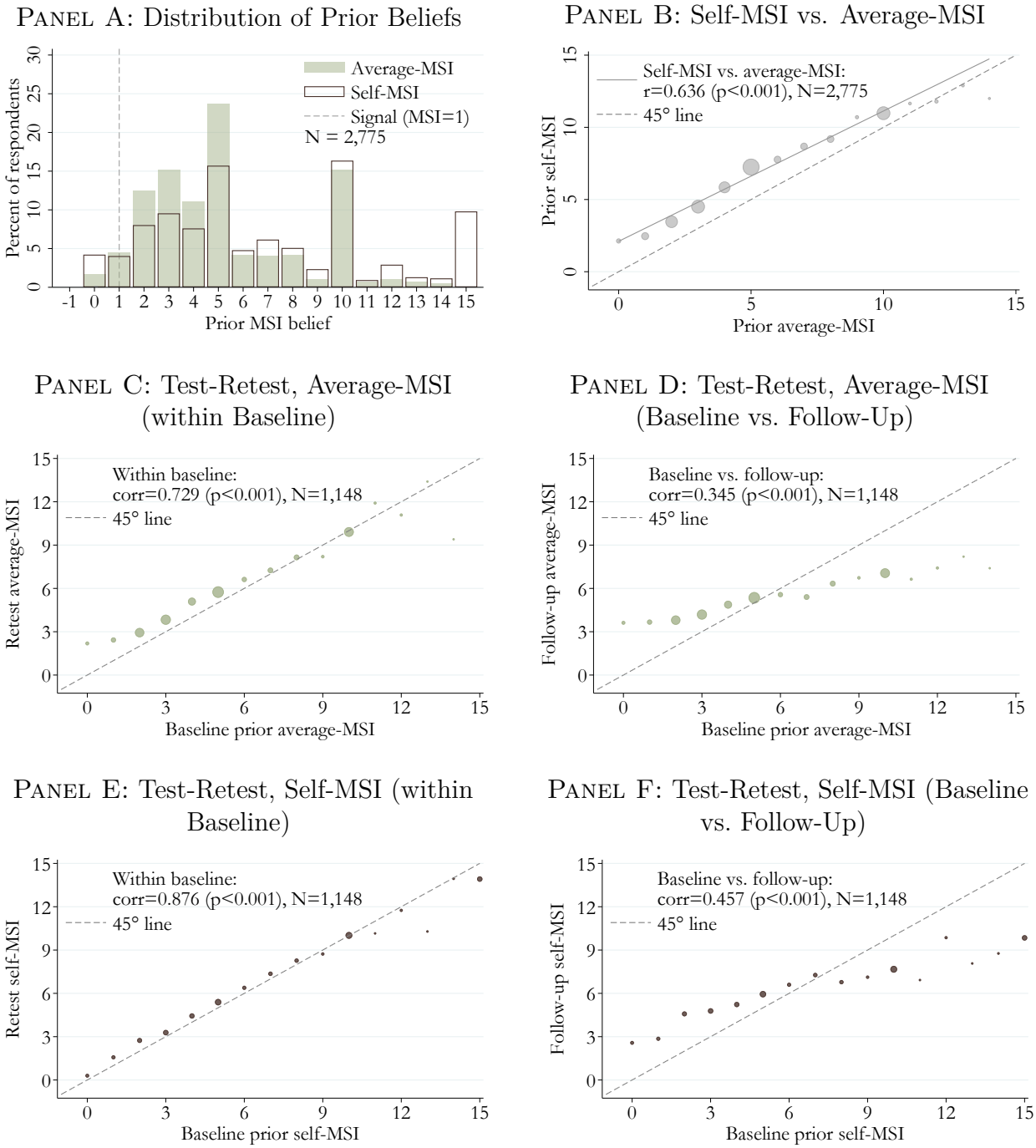
Notes: This figure shows examples of the two choice elicitation modules from a real survey participant in our main analysis sample. Panel A shows the job-choice scenario involving work hours. Panel B shows the real-world decision that was generated from information provided by the respondent, which in this case also involves extra hours worked.

Figure 3: What Real-World Decisions Weigh against Income



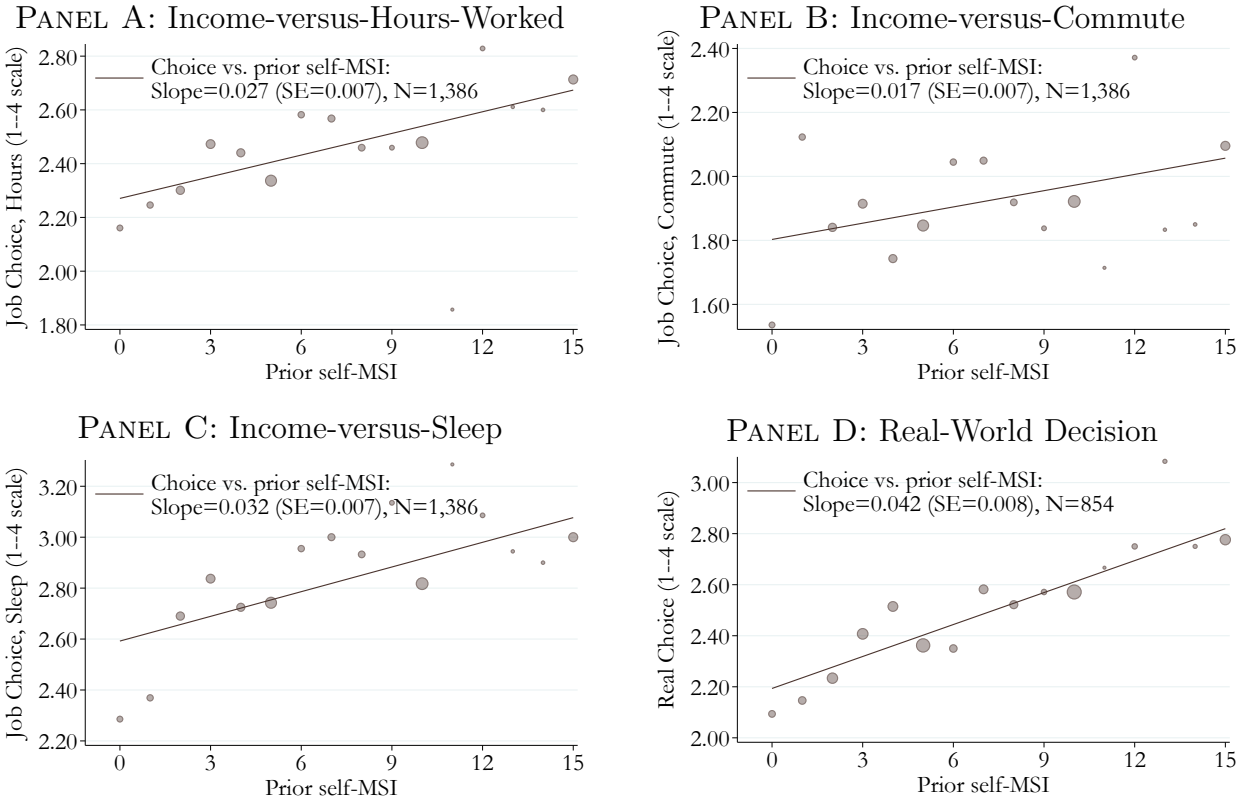
Notes: Distribution of the real-world decisions in the baseline analysis sample by the income consideration involved (Panel A) and the non-income consideration weighed against income (Panel B). For each decision, we ask a language model (gpt-5-mini) to describe each side in the respondent's own words; the descriptions are then clustered with an unsupervised topic model (BERTopic), and closely related topics are grouped into the categories shown. Appendix C.5 gives the full procedure and the complete, unaggregated topics (Tables C.3–C.4). Panel C shows the ten most common combinations of an income and a non-income category, which together account for 58.6% of the decisions with an identified theme on both sides. Panels A and B each show the nine largest categories; the remaining categories are grouped into "Other." Bars show the percentage of decisions (Panel A: 1,710; Panel B: 1,708; Panel C: 1,498 with a theme on both sides).

Figure 4: Prior Beliefs and Measurement



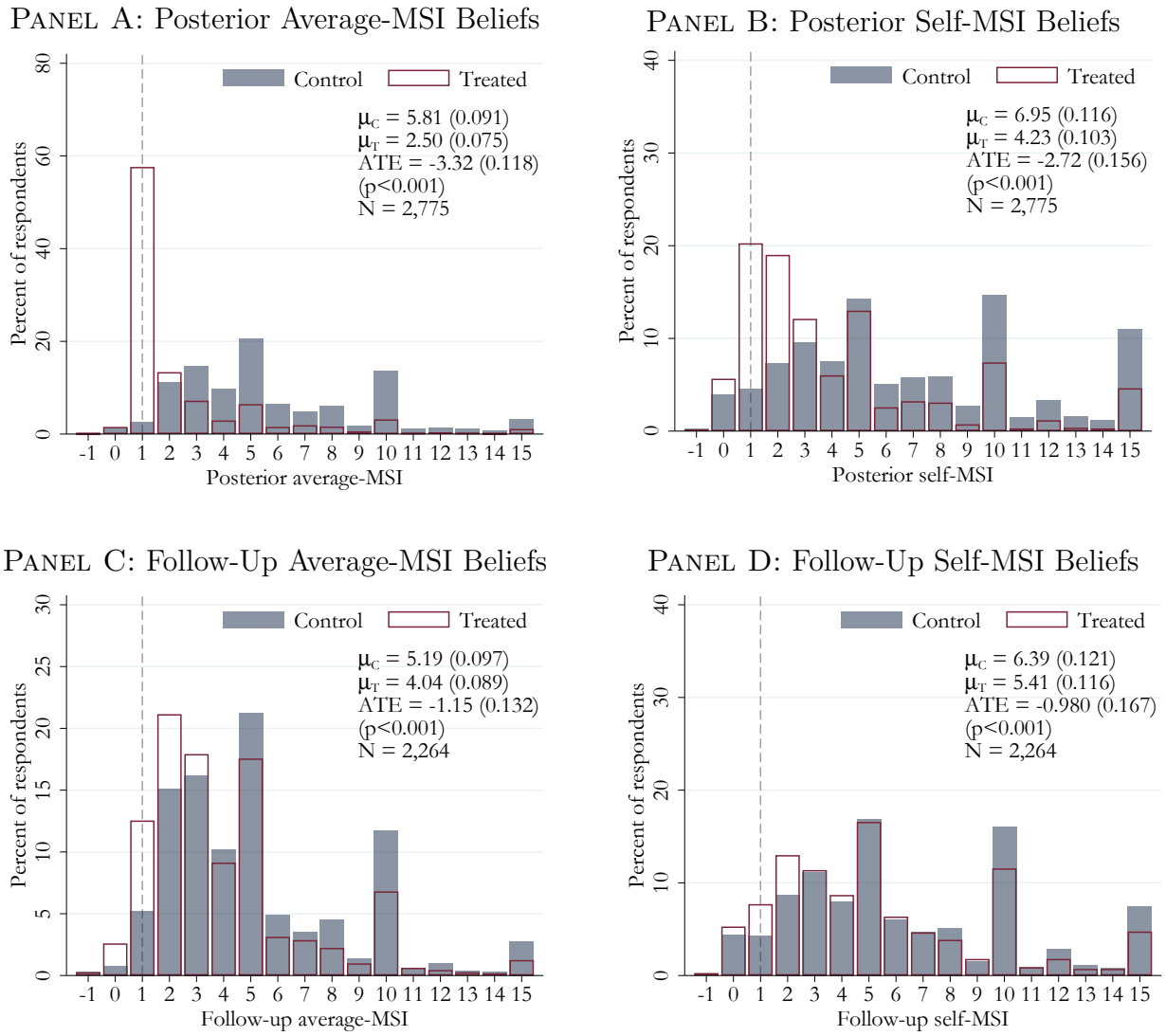
Notes: Panel A reports the distributions of prior average-MSI and self-MSI beliefs. Panel B compares prior self-MSI beliefs to prior average-MSI beliefs. Panels C–F report test-retest checks for the belief measures, based exclusively on control-arm respondents who also completed the follow-up survey. Panels C and D correspond to average-MSI beliefs and Panels E and F to self-MSI beliefs. Panels C and E compare the two baseline elicitations (within-baseline test-retest); Panels D and F compare the baseline elicitation to the same belief measured in the follow-up survey about one month later. Each legend reports the correlation coefficient and its p-value.

Figure 5: Prior Self-MSI Beliefs and Choices in the Control Arm



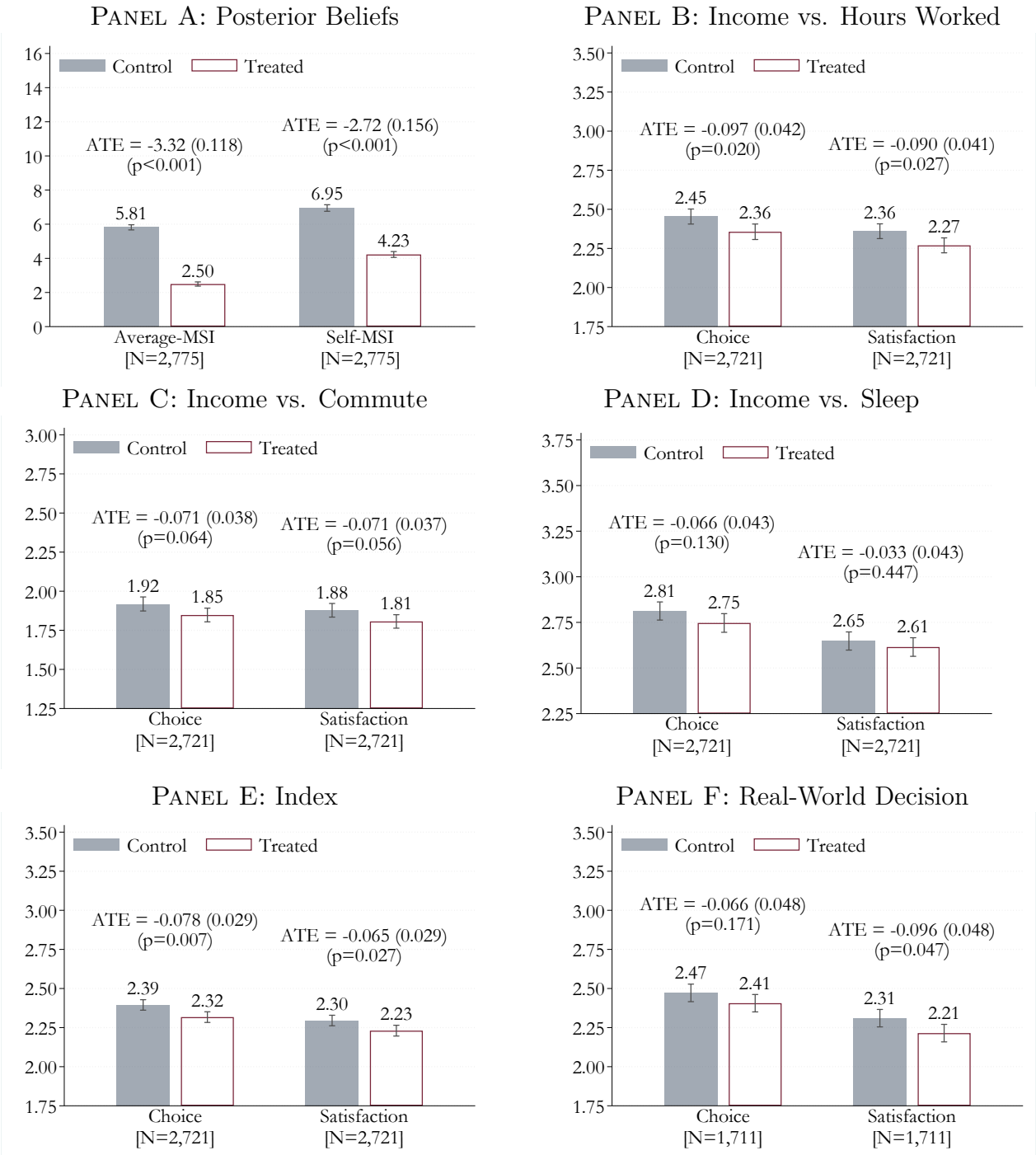
Notes: This figure reports binned scatterplots for subjects in the control arm. Panels A–C relate prior self-MSI beliefs to the three job-choice outcomes. Panel D relates prior self-MSI beliefs to the real-world decision outcome. Higher values of the choice outcomes indicate greater willingness to choose the higher-income option. The choice outcomes are on the raw 1–4 scale. Each legend reports the slope coefficient, its standard error, and the number of observations.

Figure 6: Posterior Beliefs by Treatment Status



Notes: Panels A and B report the distributions of posterior average-MSI and self-MSI beliefs, elicited immediately after treatment, by treatment status. Panel A shows average-MSI beliefs, and Panel B shows self-MSI beliefs. Panels C and D report the corresponding posterior beliefs at the one-month follow-up. The dashed vertical line marks the treatment signal (MSI = 1). Within each panel we report the control-group mean (μ_C), the treated-group mean (μ_T), and the average treatment effect (ATE = $\mu_T - \mu_C$), with robust standard errors in parentheses, the associated p-value, and the number of observations (N).

Figure 7: Average Treatment Effects of the Information Treatment



Notes: Bars are group means with 90% confidence intervals; each annotation gives the average treatment effect, robust standard error in parentheses, and p-value. In Panel A, the outcomes are the posterior beliefs about average-MSI and self-MSI, in life-satisfaction points (0–100 scale). Panels B–E are the job-choice income trade-offs against work hours, commute, sleep, and their index; Panel F is the real-world decision. These behavioral outcomes are coded on a 1–4 scale so that a higher value corresponds to a preference for the higher-income option. The number of observations is reported in brackets below each label.

Table 1: Randomization Balance and Sample Means

	Sample means			
	All	Control	Treated	p-value
	(1)	(2)	(3)	(4)
Panel A: Pre-treatment Characteristics				
Age 18-34	0.430	0.434	0.426	0.683
	(0.009)	(0.013)	(0.013)	
Age 35-54	0.408	0.401	0.415	0.458
	(0.009)	(0.013)	(0.013)	
Age 55+	0.162	0.165	0.159	0.659
	(0.007)	(0.010)	(0.010)	
Female (0/1)	0.488	0.482	0.495	0.497
	(0.009)	(0.013)	(0.014)	
White (0/1)	0.644	0.655	0.633	0.217
	(0.009)	(0.013)	(0.013)	
Edu. HS or some college	0.312	0.306	0.317	0.542
	(0.009)	(0.012)	(0.013)	
Edu. Assoc./BSc degree	0.502	0.502	0.502	0.982
	(0.009)	(0.013)	(0.014)	
Edu. Postgraduate	0.186	0.192	0.181	0.450
	(0.007)	(0.010)	(0.010)	
HH income ($1,000s)	91.0	91.2	90.8	0.840
	(1.13)	(1.63)	(1.55)	
Democrat (0/1)	0.441	0.438	0.444	0.745
	(0.009)	(0.013)	(0.013)	
Republican (0/1)	0.235	0.235	0.236	0.964
	(0.008)	(0.011)	(0.012)	
Baseline life sat (0-100)	62.1	61.9	62.4	0.569
	(0.449)	(0.637)	(0.632)	
Employed (0/1)	0.758	0.747	0.770	0.147
	(0.008)	(0.012)	(0.011)	
Prior average-MSI	5.20	5.18	5.23	0.629
	(0.057)	(0.080)	(0.081)	
Prior self-MSI	6.80	6.81	6.80	0.950
	(0.081)	(0.114)	(0.115)	
Panel B: Attrition				
Responded to Follow-up	0.823	0.821	0.826	0.728
	(0.007)	(0.010)	(0.010)	
Observations	2,775	1,413	1,362	

Notes: Main analysis sample: excludes respondents who clearly failed the video-code check or failed an end-of-survey attention check and respondents with extreme prior average-MSI (see Section 4.3). Column (1) is the pooled-sample mean; columns (2) and (3) are the control and treated means; column (4) is the two-sided p-value for the treated-minus-control difference (OLS with robust standard errors). Standard errors are reported in parentheses below each mean. The bottom row reports observation counts in each subsample. A joint test of treatment status on all balance covariates yields $F = 0.532$ (p=0.906).

Table 2: The Effect of Self-MSI Beliefs on Choices: OLS and 2SLS Estimates

	Job-Choice Scenario: Income vs. ...				
	Work	Commute	Sleep	Index	Real Dec.
	(1)	(2)	(3)	(4)	(5)
Panel A: OLS					
Choice-Framing	0.025***	0.020***	0.034***	0.026***	0.027***
	(0.005)	(0.005)	(0.005)	(0.003)	(0.006)
Satisfaction-Framing	0.026***	0.020***	0.031***	0.026***	0.032***
	(0.005)	(0.004)	(0.005)	(0.004)	(0.006)
Panel B: 2SLS					
Choice-Framing	0.035**	0.026*	0.024	0.028***	0.024
	(0.015)	(0.014)	(0.016)	(0.010)	(0.017)
Satisfaction-Framing	0.033**	0.026*	0.012	0.024**	0.035**
	(0.015)	(0.014)	(0.016)	(0.011)	(0.017)
Panel C: 2SLS at Follow-Up					
Choice-Framing	0.026	-0.004	0.034	0.019	-0.006
	(0.046)	(0.042)	(0.046)	(0.031)	(0.076)
Observations					
Baseline	2,721	2,721	2,721	2,721	1,711
Follow-Up	2,225	2,225	2,225	2,225	1,384
Dep. Var. Mean (Control Group)					
Choice-Framing	2.45	1.92	2.81	2.39	2.47
Satisfaction-Framing	2.36	1.88	2.65	2.30	2.31
Dep. Var. SD (Control Group)					
Choice-Framing	1.09	1.02	1.12	0.76	1.00
Satisfaction-Framing	1.07	0.99	1.13	0.76	0.99
Kleibergen-Paap rk Wald F statistic					
Baseline, Choice-Framing	307	307	307	307	198
Baseline, Satisfaction-Framing	307	307	307	307	198
Follow-up, Choice-Framing	34.9	34.9	34.9	34.9	20.6

Notes: Each column is an outcome on a 1–4 scale, coded so that 1 indicates definitely choosing the lower-income option and 4 indicates definitely choosing the higher-income option: the job-choice income trade-off against work hours (1), commute (2), sleep (3), their index (4), and the real-world decision (5). At follow-up (Panel C, column (5)), the classification had to be harmonized—see Appendix C.6 for more details. Entries are coefficients on posterior self-MSI beliefs, with robust standard errors in parentheses. Panel A is OLS; Panel B is 2SLS instrumenting posterior self-MSI with the randomized treatment; Panel C repeats the 2SLS at the one-month follow-up, with both the outcome and posterior self-MSI measured at follow-up. Each panel reports rows for the choice and satisfaction framings (Panel C, choice framing only). Bottom rows give the sample size, the control-group dependent-variable mean and SD, and the Kleibergen-Paap *rk* Wald F-statistic for instrument strength. Sample: main analysis sample (Panel C: follow-up sample). *p<0.10, **p<0.05, ***p<0.01.