THREE ESSAYS ON APPLIED MICROECONOMICS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Daniel Beshears Reeves

August 2018

THREE ESSAYS ON APPLIED MICROECONOMICS

Daniel Beshears Reeves, Ph.D.

Cornell University 2018

This dissertation consists of three essays at the intersection of behavioral economics and public finance. Two essays focus on the topic of income taxes, in particular, income-tax salience and income-tax perceptions, and how these differ for federal vs. state income taxes. These perceptions have potential implications for optimal policy design and labor-supply decisions. The third essay uses a novel approach to examine potential nonresponse bias in large government surveys that are a frequent source of economics data.

In Chapter 1, I examine income-tax salience. Income taxes distort labor market outcomes, sometimes as a primary goal (as in the earned income tax credit) and sometimes as a secondary consequence (as a by-product of raising revenue). However, for an individual's decisions to be influenced by a tax, he needs to be incorporating it in his decision-making. If the source of the tax—e.g., the state or federal government—affects the consumer's awareness of the tax in his decision-making, then different income taxes might have different labor-market effects. This chapter provides experimental survey evidence that state taxes are less salient than federal taxes. Reduced-form estimates show a large difference between state and federal taxes. Direct estimation of salience parameters in a preferred subsample indicates that federal income taxes are close to fully salient, while state income taxes are substantially and significantly less salient. The chapter also explores evidence for individual-level explanations of salience and tax knowledge.

In Chapter 2, I examine income-tax perceptions. Income taxes are complicated

and potentially confusing to consumers. Individuals may rely on heuristics or make systematic mistakes in their perceptions of their taxes. This chapter uses experimental survey data to add to a literature seeking to understand federal income-tax perceptions and adds novel evidence on state income-tax perceptions. I find broad agreement in perceptions between state and federal income taxes. Generally, individuals perceive too much progressivity in both federal and state tax schedules. Estimating a model of federal tax perceptions yields results consistent with a more complicated mental model than suggested in previous literature.

In Chapter 3, we examine nonresponse bias in large government surveys. Economists are often interested in measuring outcomes such as unemployment, labor force participation, obesity, and household expenditures. In this chapter we study the sources of the relevant official statistics—the Current Population Survey (CPS), the Behavioral Risk Factor Surveillance System (BRFSS), and the Consumer Expenditure Survey (CEX)—and find that measurements of these outcomes depend on whether we look at easy vs. difficult-to-reach respondents. These results empirically substantiate the theoretical warning against making population-wide estimates from surveys with low response rates.

BIOGRAPHICAL SKETCH

Daniel Reeves was born and raised in Purlear, North Carolina. There he graduated from West Wilkes High School. He attended the University of North Carolina at Chapel Hill, graduating with majors in economics and political science. After college, he worked as a Research Associate at the Research Triangle Institute in their Behavioral Health Economics Program. He then returned to school to pursue a Ph.D. in Economics at Cornell University. He is joining Capital One in McLean, VA as a Principal Associate.

For the endless love, support, and guidance they have provided me. In honor of Betty Beshears, and in memory of Daniel "D.F." Beshears Jr., Mary Kate Reeves, and Glenn "Jack" Reeves.

ACKNOWLEDGEMENTS

I would like to begin by thanking my committee, all of whom deserve more thanks than space will allow. Thanks to my advisor, Ted O'Donoghue, for giving so freely of his time, energy, knowledge, and support across many years (and across many more drafts and meetings). Without his assistance, this work would never have been completed. I would also like to thank Ori Heffetz for his guidance and for taking me on as a coauthor, which was an incredibly valuable learning experience. Finally, thanks also to Tatiana Homonoff, whose insightful comments, support, and encouragement were invaluable to the process.

While my committee members were the ones to guide me along the final step of my educational journey, I also benefited from an amazing cast of educators along the way. My sincere thanks to each of those wonderful teachers and librarians.

My very first and very best teacher was my mother, Debbie Reeves. Without a doubt, the countless hours she spent reading to me throughout my childhood instilled a lifetime love of learning in me. I can trace back every accomplishment to the unconditional support both her and my father, Doug Reeves, have given me in every endeavor. My brothers, Jordan and Jackson, are pretty cool too.

Also, thanks to the many wonderful and talented people I have met at Cornell. In particular, to my first-year study group who got me through Q's. Vesall and Sally always provided a calm harbor amidst any storm. James, Michiel, and Anne reminded me grad school can have a social side. And, finally, thanks to Miriam for being always being Miriam.

Finally, thanks most of all to my constant companion of (almost) the whole process, my travel buddy, my one-woman support system, my cooking partner, my motivator, and my often-times counselor: Lin Xu. She has provided me more assistance, in more ways, than anyone has ever deserved.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

ARE ALL INCOME TAXES REALLY THE SAME? AN EXPERIMENTAL INVESTIGATION INTO STATE AND FEDERAL INCOME-TAX SALIENCE

1.1 Introduction

Public-finance research typically assumes that individuals know about and react appropriately to all relevant taxes when they make decisions about which goods to purchase and how much labor to supply to the market (Mirrlees, 1971; Diamond, 1980). However, recent evidence demonstrates that this classical assumption may not always match how individuals actually make decisions. Indeed, recent studies show that characteristics of both the individual and the tax can affect the way in which people respond to a tax (Chetty et al., 2009; Taubinksy and Rees-Jones, 2017).

This paper studies whether people react differently to state and federal income taxes. The source of an income tax should be irrelevant to labor-supply decisions in a classical labor model. However, state and federal income taxes differ substantially in many domains: their discussion in national media, debate in political campaigns, and in their average magnitudes. Any or all of these factors might create a gap in salience between the two taxes. The difference in salience could lead to differences in how individuals respond to each tax, despite the classical assumption of equivalence. This paper examines individuals' perceptions of taxes using an experimental, incentivized elicitation of the individuals' reports of various income tax rates. I find evidence indicating individuals find state income taxes substantially less salient than federal income taxes.

The experiment has three main goals: (1) measure the relative salience of state and federal income taxes, (2) measure relative knowledge and understanding of state and federal income taxes, and (3) examine individual and structural determinants of income-tax knowledge and salience. The online experiment uses the Amazon MTurk platform to collect data from over 3,000 U.S. workers to elicit their perceived total, federal, and state marginal income-tax rates. The MTurk sample presents a relatively rare group of workers who are making intensive margin labor-supply decisions (choosing exactly how many MTurk tasks to perform for pay). The experiment manipulates salience of the different types of taxes through three conditions to test for salience differences. The two treatment conditions draw the respondents' attention to either state or federal taxes, depending on condition, by asking them a series of five questions about what either the state or federal rate would be at different incomes, and then asking the respondent to use one of these rates in a simple calculation. The state treatment gets the state questions, while the federal treatment gets the federal questions. In the control condition respondents are not asked about specific income taxes, only total income taxes. The initial effect of these treatments is measured by the change in reported total tax rates.

I find several pieces of evidence consistent with a difference in the salience of state and federal taxes. Respondents self-report the first piece of evidence when asked what types of taxes they were including in their total tax-rate answers. Overall, 55% of respondents include both state and federal income taxes in their answers, while 35% report only including federal income taxes. Breaking down the results by treatment group makes the differences in state and federal tax salience much more stark. Participants in the control group include both state and federal income taxes only 35% of the time, while including only federal income taxes 54%

of the time. Comparatively, these numbers dramatically flip in the state treatment: respondents include both taxes 72% of the time, while only 17% of people report including only federal taxes. Results in the federal treatment, in both cases, lie between the state and control conditions and will be discussed in detail in Section 1.4. These self-reports indicate that respondents tend to ignore state taxes unless something, such as the treatment, makes them more salient.

Regression analysis provides further evidence of a difference in salience. Respondents are asked their perceived total tax rate before and after treatment. Respondents in the state and federal treatment conditions have higher increases in their reported total tax rate from pre- to post-treatment compared to the control group. Treated groups' reports increase by approximately 2.5 percentage points between rounds (both significant at the 1% level); the control group has no significant change. Notably, the change in the treated groups' reports is of the same magnitude as the median state income tax rate in the sample (4.5%). This result suggests that many respondents were not initially including state taxes in their estimates in round 1, but do so once the treatment makes state taxes salient. Additionally, regressions show that individuals' reported state rate is incorporated less in their total rate answers than is their reported federal rates (i.e. one percentage point of perceived state taxes increases their total less than one percentage point of federal taxes). Specifically, in the absence of treatment, a one percentage-point increase in a respondent's reported state rate increases their total report by 0.1 percentage points, while an equivalent change in reported federal rate yields a 0.7 percentage-point increase. The state treatment substantially increases the effect of reported state rates, giving it a total impact of 0.2 percentage points (statistically significant at the 1% level). These estimates reflect a much stronger impact of an individual's federal rate on their beliefs about their total rate relative to their state

rate.

Finally, this paper builds and estimates a simple model of tax salience. In this model, individuals' perceptions of their state and federal tax rates are functions of their true rates, but individuals may systematically over- or under-estimate a given tax rate (empirically, respondents systematically over-estimate the state and federal tax rates that they face). Total tax perception is a weighted combination of specific tax rates, where the weight is a salience parameter specific to the tax. If a tax is less salient it will have a lower salience parameter and, correspondingly, have less effect on perceived total tax rates. Estimating this model yields salience parameters consistent with state taxes being less salient than federal taxes. The state salience parameter is 0.8 compared with a federal salience parameter of 1.1. (A standard, rational agent would have full salience for both taxes: a salience parameter of 1.)

This paper adds to a literature on the importance of salience in tax decisions. Chetty et al. (2009) study the differences between state sales and excise taxes using both experimental and field data. They find that individuals react less to taxes applied at the register (sales taxes) than to taxes included in the list price of the good itself (excise taxes). The authors also highlight that salience affects optimal-tax design, as less salient taxes distort consumer behavior by less. Similarly, Finkelstein (2009) finds that tolls that are paid in a less salient manner (electronically, instead of in cash) create less response in consumer behavior, leading to higher toll rates in areas with electronic payment. Recent contributions to this literature note that characteristics of the consumer are also an important predictor of salience, as low-income individuals pay attention and respond more fully to low-salience sales taxes than high-income individuals (Goldin and Homonoff, 2013). This paper's

results further illustrate the importance of considering salience in tax design since differential salience between groups has distributional welfare effects. This point is reiterated in experimental evidence on sales taxes by Taubinksy and Rees-Jones (2017), who find that there are heterogeneous types in the population with different attentiveness to less salient sales taxes; this difference in attentiveness persists even at artificially-created extreme tax rates.

A related literature explores individuals' knowledge of and (mis)perceptions of income taxes. Survey analysis in Gideon (2017) provides evidence that individuals do not correctly perceive the progressive nature of federal income taxes, underestimating its progressivity, by overestimating average tax rates and underestimating marginal tax rates. Other work shows that the complicated nature of the federal tax structure causes people to incorrectly react to tax credits (Feldman et al., 2016; Miller and Mumford, 2015). While economic theory predicts that individuals will react to changing tax schedules by bunching at discontinuities, field evidence shows that many taxpayers do not react to these changes (Saez, 2010). Furthermore, experimental evidence from Rees-Jones and Taubinsky (2016) shows the classical perception bias of ironing (using average rates instead of marginal rates) in individuals' reports of the tax schedule.

By presenting results from an experiment, this paper will contribute novel evidence in the domain of state income-tax salience and knowledge, an area that has been relatively unstudied (research has largely focused on sales and excise taxes). State income taxes are important as both a policy lever and revenue source for state policymakers. In 2014, state income taxes made up approximately one-fifth of state government revenues and was the second largest tax source for state governments, after only sales taxes (TPC, 2017). This number represents

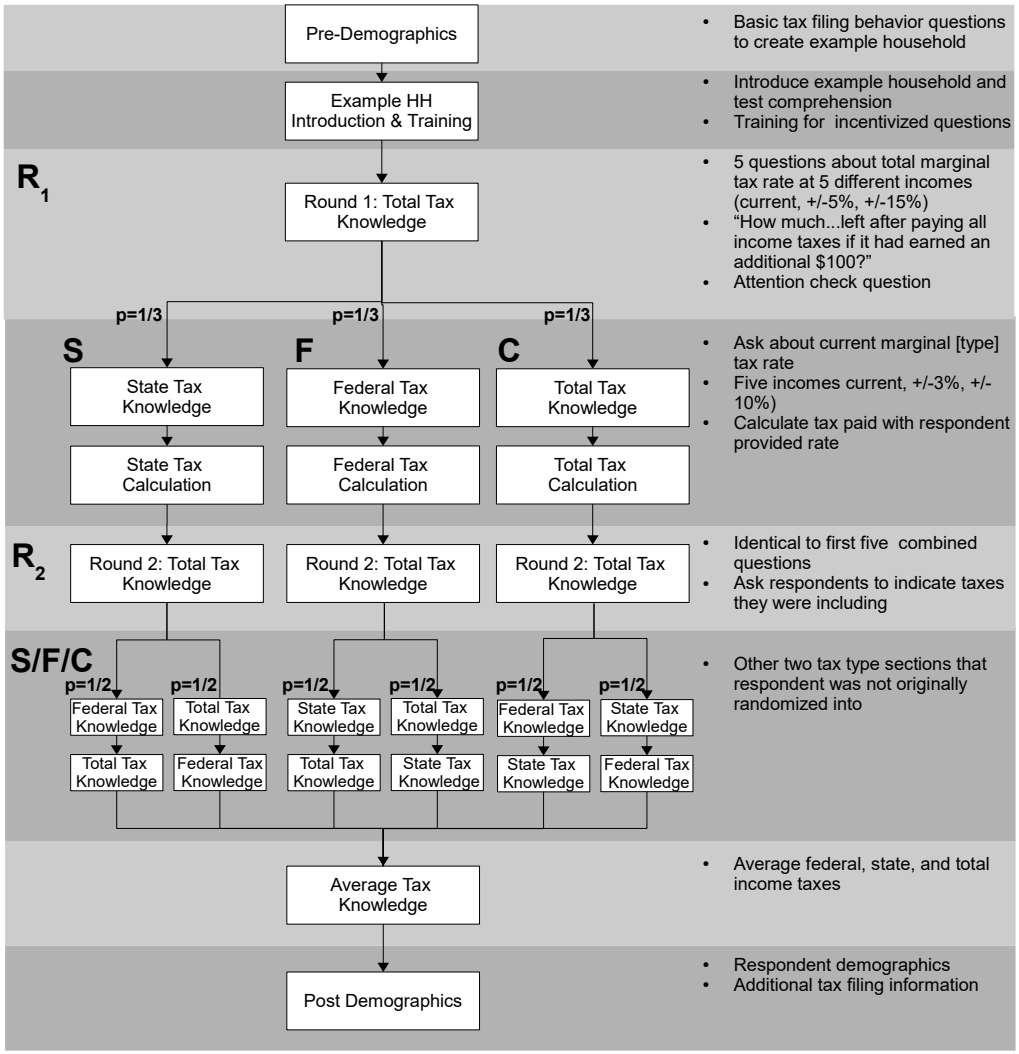
over $300 billion in annual state income taxes paid by consumers and highlights the economic importance of understanding how consumers do or do not correctly perceive and react to these state taxes.

Much like sales and excise taxes, different sources of income taxes should have the same final effect on consumer behavior through their effect on the final price. However, unlike sales and excise taxes, there is not an immediately identifiable mechanism (e.g. posted versus register price) through which the salience operates for income taxes. Therefore, this paper highlights a new potential channel for studying salience in public finance and a new source for identifying factors influencing salience. This paper begins this search for mechanisms through studying a variety of individual and policy characteristics to determine what might effect salience. While this paper's initial findings on this front are inconclusive, the consistent evidence of differential salience indicates that this is a potentially fruitful area for future research.

1.2 Experiment

The experiment aims to measure individuals' perceived marginal tax rates and manipulate the salience between state and federal income taxes. Figure 1.1 illustrates the experiment's overall design. The experiment's core is divided into five primary sections, each containing five questions. These questions ask about the marginal tax rate at a given income for an example household, similar to the respondent's own. The primary question gives the respondent an initial income level and asks "How much more would the [example household] have, after paying **all income taxes** on the additional earnings, if it had made an additional $100?", where the

Figure 1.1: Experiment design



Notes: This figure illustrates the experimental design. All participants begin with the same three sections before being randomized to one of two treatment groups (State or Federal) or the control group. By the end of the experiment, all participants receive each of the three main sections $\mathbf{R_1}$, **S**, **F**, **C**, and $\mathbf{R_2}$. However, randomization controls the order that they receive **S**, **F**, and **C**.

specific type of tax is then varied by section.

Three sections use this exact question to ask about the combined total marginal income tax rate faced by the household. In the figure these are sections Round 1 ($\mathbf{R_1}$), Round 2 ($\mathbf{R_2}$), and Combined ($\mathbf{C}$). Another section focuses on federal income taxes, labeled $\mathbf{F}$ (Federal) in the figure. The final section, State ($\mathbf{S}$), asks about state income taxes.

The treatments manipulate the salience of state and federal income taxes by changing the order in which the respondents are asked for sections $\mathbf{S}$ and $\mathbf{F}$. Asking about a specific tax rate should make a less salient tax more salient to the individual. However, if the tax was already fully salient, then asking about it should not change its salience. Asking the respondent about no specific taxes in section $\mathbf{C}$ serves as a control. Each section follows the same general format: marginal tax rates are elicited at five income levels, centered on the respondent's household income. The sections differ only in which specific type of income tax is asked about and at what income levels. Salience is initially measured by seeing how the respondents' answers in $\mathbf{R_2}$ change after being asked specifically about either state or federal income taxes, depending on their treatment.

Respondents are recruited using Amazon's Mechanical Turk (MTurk) platform. Respondents need to have completed at least 100 previous assignments on the platform with an approval rate of at least 95% to be eligible to participate in the experiment. Only residents of the U.S., as determined by Amazon's address verification, are eligible. Respondents can complete the experiment only once.[1] Primary data collection occurred during the tax filing season for 2016: January 27th, 2017 through April 3rd, 2017. An additional data collection period aiming

[1]Respondents to early pilot versions of the experiment are not eligible to participate in the final version of the experiment (as verified by the worker's MTurk ID, assigned by Amazon).

to collect at least 300 respondents from each of California and New York (the two largest states with progressive tax schedules) continued until May 5th, 2017.

In many settings, including this one, there are concerns about the external validity of findings from online samples. For this specific question, however, online workers provide an interesting sample of individuals who have control over their labor supply on the intensive margin: they choose how many online tasks to complete. Unlike most workers, who tend to have less intensive margin flexibility, the marginal tax rates may be particularly relevant for this sample. (Although, given potential for effort and measurement concerns, discussed later, I do not rely on this claim and, instead, only note that this sample's intensive margin flexibility makes them an interesting group to study.)

The survey begins with a brief series of questions to determine the respondent's eligibility to participate and to elicit basic information on their household's tax filing situation for 2016. The respondent is asked to provide the household's state, income, and basic tax filing information (number of dependents, filing type, and disability status of household members). If the respondent knows and provides the relevant information, the respondent is then introduced to the example household, which serves as the basis for subsequent tax questions. The example household is customized to be similar to the respondent in key features (household income, state of residence, filing type and number of dependents), but has a (potentially) simplified tax situation. The primary simplifications are that the example household only has salary income, uses only standard deductions, and lives in a location without a city or county income tax.

Respondents must then demonstrate understanding of the example household through a brief quiz on its tax-related characteristics. Respondents that pass

the quiz are given two sample questions, along with feedback, illustrating how to answer the tax questions that make up the bulk of the experiment. These training questions differ from the real questions in that a randomly chosen tax rate is provided and there is never a mention of a specific type of income tax (unlike the sections **S** or **F**, but equivalent to sections $\mathbf{R_1}$, $\mathbf{R_2}$, and **C**).

Respondents are paid a participation fee of \$1 for completing the entire survey.[2] Each respondent can then earn up to an additional \$0.50 depending on their accuracy. Respondents are informed that one of the post-training questions will be randomly selected for payment. Payment is determined using a quadratic loss formula and how close the question's answer is to the true answer.[3] The training questions include an explanation of the payment scheme and an example illustrating how wrong answers lead to a lower bonus payment. The correct tax rates were determined using Taxsim9.[4]

After training, the respondent is asked to report on how much the household would pay in marginal income taxes on \$100 in additional income at five different income levels as part of $\mathbf{R_1}$. The five income levels are the respondent's current income and that income plus and minus both five and fifteen percent. After $\mathbf{R_1}$, the respondent is randomized to one of three sections (**S**, **F**, or **C**) which determines the type of taxes the questions asks about. In the first condition, state treatment, the respondent is asked about state income taxes at five different in-

[2]This completion fee was initially \$0.50 but was increased to \$1 to recruit the target primary sample of 3,000 respondents.

[3]The formula used was $Payment = \$0.50 - 10 * (True\ tax\ rate - Implied\ tax\ rate\ from\ answer)^2 * 0.50$.

[4]Due to the timing of the experiment, designed to occur in the tax filing season for 2016, the state rate estimates in Taxsim were not yet available, so state rates from 2015 were used. Prior to making this decision, pilot data was used to back-test this strategy for the previous 10 years of state rates changes and the effect on payment was shown to be negligible (with zero effect in the overwhelming majority of cases). 2015 rates are also being used as the actual rates in this analysis.

come levels (presented in a random order) in section **S**. These five income levels are also centered on the respondent's income, but the other four levels are plus and minus three and ten percent. The respondent is then asked to use one of their tax rate answers in a calculation to determine how much the household would pay in state taxes on an additional $600 of income. The other two treatment groups only differ in the type of income tax the respondent was asked about. In the second treatment group, the respondent is asked about federal income taxes in **F**. While the third group, the control, receives **C**, and the respondent is again asked about all income taxes (with no specific types mentioned).

After the treatment, all respondents proceed to $\mathbf{R_2}$ to receive the same five questions about all income taxes that they had previously answered in $\mathbf{R_1}$. After answering the tax-rate questions, the respondent reports on what types of taxes they included in their answers for $\mathbf{R_2}$. The respondent then receives the two sets (in random order) of specific tax questions (**S**, **F**, or **C**) that they did not receive as an initial treatment between $\mathbf{R_1}$ and $\mathbf{R_2}$. All respondents then report the average state, federal, and total taxes the example household would pay at the respondents' current income level. Finally, the respondents are asked a series of more detailed demographic and tax-filing questions.[5]

A total of 3,160 respondents complete at least $\mathbf{R_2}$ and pass the attention check embedded in the first round of questions.[6] To eliminate individuals who seem to be misunderstanding the questions or not trying to answer the survey correctly, I focus the primary analysis on the 2,326 respondents who are the most accurate on the $\mathbf{R_1}$ questions (i.e. the most accurate before any treatment). These respondents are identified by using an error index of the average absolute error on the five questions

[5]See appendix A.2 for full survey text.

[6]The attention check asks the respondent to enter the name of the example household instead of a tax rate and occurs after the five other tax-rate questions in $\mathbf{R_1}$.

Table 1.1: Demographics

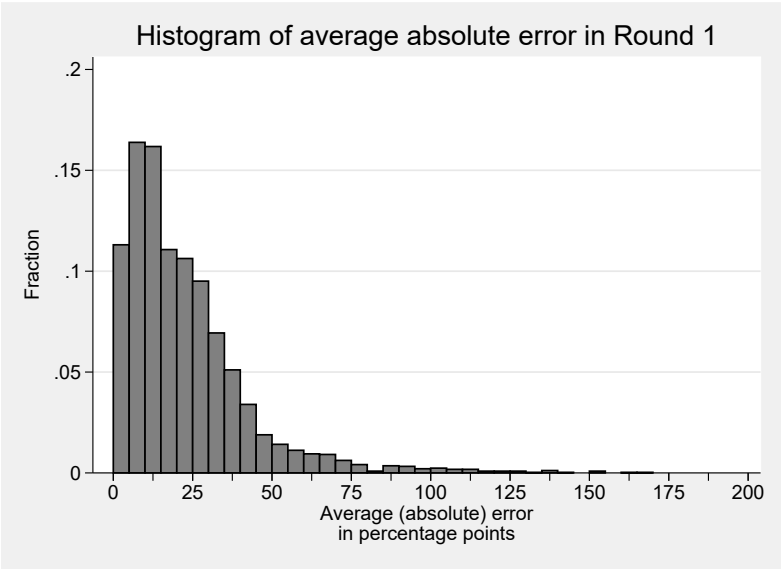
	Analysis Sample	Full sample
Female	0.568***	0.587
Age: 18–29	0.282	0.269
30–39	0.339	0.329
40–49	0.181	0.172
50–59	0.112	0.106
60 and up	0.058	0.054
Taxes prepared by: Self, by hand	0.057	0.056
Self, by software	0.613**	0.578
Other HH member	0.075	0.076
Tax professional	0.227	0.220
H.H. Income: $0– $29,999	0.216***	0.240
$30,000– $49,999	0.216	0.224
$50,000– $84,999	0.304***	0.289
$85,000 or more	0.264***	0.247
Education: High school, or less	0.060***	0.070
Some college or tech. school	0.286***	0.298
College	0.425***	0.390
Graduate or professional degree	0.216***	0.188
Income source: Salaries, wages or tips	0.847**	0.839
Self-employment income	0.071***	0.078
Government transfers	0.032	0.033
N	2,326	3,039

Notes: The right column contains the sample demographics from the individuals that completed the survey and passed the attention check embedded in $\mathbf{R_1}$. The left column is the demographics from the primary analysis sample which, in addition to finishing the survey and passing the attention check, had less the 75th percentile of total error in $\mathbf{R_1}$. Stars indicate statistically significant difference between analysis sample and excluded sample. *** p<0.01, ** p<0.05, * p<0.1.

in $\mathbf{R_1}$. Figure 1.2 presents a histogram this index. The median error per question is approximately 17 percentage points. Respondents are classified as being part of the "low error" analysis sample if they have less than the 75th percentile on this metric; corresponding to a cutoff of less than 30 percentage points of average error.

The left column of Table 1.1 contains the demographics of the primary analysis sample. The analysis sample skews female (57%). The bulk of the sample is relatively young with almost two-thirds of the sample under the age of fifty. There is a wide distribution of household income in the sample: 22% earn less than $30,000 and another 26% earn more than $85,000 a year. This wide income range

Figure 1.2: Average absolute error in round 1 "Total tax" questions.



Notes: The average error, in percentage point terms, of the five total income tax questions in R_1.

is useful for generating a variety of income taxes faced, as the federal income tax system (and many state income tax systems) are progressive: higher incomes lead to higher tax rates. In general, the sample is extremely highly educated, less than 10% of the sample did not participate in at least some post-secondary education. Usefully, two-thirds of the respondents are responsible for completing their household's taxes, either by hand (6%) or by using tax-preparation software (61%). For comparison purposes, the right column presents the demographics of the full potential sample (those who passed the attention check and completed $\mathbf{R_2}$). In general, the full sample is somewhat less likely to prepare their own taxes, is slightly less educated, and has a lower household income.

1.3 Theoretical and Empirical Framework

1.3.1 Theoretical Framework

The experiment is motivated by the following theoretical model: Individual i faces a true total marginal tax rate ($\tau^*_{T_i}$) for labor income that is the sum of his federal, state, and payroll marginal tax rates ($\tau^*_{F_i}$, $\tau^*_{S_i}$, and $\tau^*_{P_i}$ respectively):

$$\tau^*_{T_i} = \tau^*_{F_i} + \tau^*_{S_i} + \tau^*_{P_i} \tag{1.1}$$

In practice, there is not a sufficient amount of useful variation in the payroll tax faced by households in my sample since only very high-earning households experience changes in their payroll tax rate. Additionally, I do not separately observe participants' perceptions of payroll taxes, therefore I limit my modeling focus to the individual's state and federal taxes.

For each specific tax type L, the individual has a perception (τ_{L_i}) that may be different from their true tax rate ($\tau^*_{L_i}$). Their perceived federal and state tax rates are then:

$$\tau_{F_i} = \beta_{F,0} + \beta_{F,1}\tau^*_{F_i} + \delta_{F_i} \tag{1.2}$$

$$\tau_{S_i} = \beta_{S,0} + \beta_{S,1}\tau^*_{S_i} + \delta_{S_i} \tag{1.3}$$

$\beta_{L,0}$ captures the degree to which all individuals systematically either overestimate (if $\beta_L > 0$) or underestimate ($\beta_L < 0$) the specific tax and δ_{L_i} represents the individual-specific error in perception around the true specific tax rate. If individuals are fully rational then $\beta_{S,0} = \beta_{F,0} = 0$ as individuals will make no systematic mistakes in perceiving their taxes. $\beta_{L,1}$ represents the extent to which the true tax

rate is reflected in the individual's perception.

The individual's perception of their total marginal tax rate is a weighted combination of their individual tax rates, where each specific tax rate is weighted by a salience parameter (θ_L):

$$\tau_{T_i} = \beta_{T,0} + \theta_F \tau_{F_i} + \theta_S \tau_{S_i} \tag{1.4}$$

For fully knowledgeable, rational individuals $\beta_{T,0}$ will equal τ_P^* as they correctly include their payroll taxes in their total tax perception. θ_F and θ_S will also both equal 1 as each specific tax perception is fully incorporated into their perception of the total tax rate. However, if individuals find some taxes less salient, then perceptions of that tax will not be fully reflected in their total perceived tax rate and θ_L may be less than 1.

Each individual i's perception of tax rate L is elicited at trial j with error (ϵ_{ij}), so that the observed perceptions of tax rate L ($\hat{\tau_{L_i}}$) are:

$$\hat{\tau_{F_i}} = \beta_{F,0} + \beta_{F,1}\tau_{F_i}^* + \delta_{F_i} + \epsilon_{ij} \tag{1.5}$$

$$\hat{\tau_{S_i}} = \beta_{S,0} + \beta_{S,1}\tau_{S_i}^* + \delta_{S_i} + \epsilon_{ij} \tag{1.6}$$

$$\hat{\tau_{T_i}} = \beta_T + \theta_F \tau_{F_i} + \theta_S \tau_{S_i} + \epsilon_{ij} \tag{1.7}$$

Substitution yields that the individual's observed perception of total tax rate will be:

$$\hat{\tau_{T_i}} = \beta_T + \theta_F(\beta_{F,0} + \beta_{F,1}\tau_{F_i}^* + \delta_{F_i}) + \theta_S(\beta_{S,0} + \beta_{S,1}\tau_{S_i}^* + \delta_{S_i}) + \epsilon_{ij} \tag{1.8}$$

The experiment manipulates the salience of the taxes for individuals for whom the taxes are not salient. Asking about specific taxes likely increases their salience to the individual, so that being asked to identify their state tax rate will increase its salience when the individual later gives an total tax rate. Therefore the salience of a tax (θ_L^H) will depend on the treatment history, H, of the individual. This predicts that $\theta_S^S \geq \theta_S^C$ and $\theta_S^S \geq \theta_S$: state taxes will be more salient for those in the state treatment than in the control and will be more salient for those in the state treatment than the untreated, respectively.

While not an ex-ante prediction, it also became clear during pilot testing the experiment that thinking about any given specific rate increases the salience of all the other rates, as it may remind the individual that other rates exist. Thus state tax salience is higher in the federal treatment than it is initially or in the control treatment: $\theta_S^F \geq \theta_S$ and $\theta_S^F \geq \theta_S^C$. The salience-increasing effect of asking about a specific tax rate is likely to be strongest on that specific tax (relative to its effect on the salience of other taxes) so that $\theta_S^S > \theta_S^F$.

1.3.2 Empirical Framework

I focus on three primary regressions to get a descriptive estimate of the effect of salience. The first examines the average impact of treatment on how answers to the total tax questions changed between round 1 and round 2:

$$R_2 - R_1 = \beta_0 + \beta_1 I_S + \beta_2 I_F \tag{1.9}$$

The dependent variable is the total tax rate implied by the round-1 answer (R_1) subtracted from the round-2 tax rate (R_2). A positive change indicates that the

reported total tax rate went up between rounds, while a negative change indicates that the tax rate went down. The two independent variables are indicators for the treatment group the respondent is assigned to, either state (I_S) or federal (I_F), while the omitted category is the control group. If, as hypothesized, state taxes are not initially salient to individuals and being randomized to the state treatment group increases the salience of state taxes between the rounds then β_1 should be positive.

Regression 1.10, examines the degree to which respondents' round-2 reported total-tax rates reflect their reported state and federal tax rates, and the degree to which treatment affects this:

$$R_2 = \beta_0 + \beta_1 I_S + \beta_2 I_F + \beta_3 \hat{\tau_S} + \beta_4 I_S \hat{\tau_S} + \beta_4 I_F \hat{\tau_S} + \beta_5 \hat{\tau_F} + \beta_6 I_S \hat{\tau_F} + \beta_7 I_F \hat{\tau_F} \quad (1.10)$$

In this equation $\hat{\tau_S}$ and $\hat{\tau_F}$ are the respondents' reported state and federal tax rates, respectively. Once again, I_S and I_F are indicators for assignment to the state or federal treatment groups. In this regression, treatment assignment is interacted separately with the respondent's reported rate for each type of tax. If individuals are perfectly rational, i.e. both taxes are fully salient, then β_3 and β_5 would be the only statistically significant coefficients and their value would be expected to be 1: all respondents' answers would include their reported state and federal income tax rates. The degree to which these coefficients are initially less than one indicates the relative salience of different taxes and the interaction terms indicate how much the treatments impacted salience.

Equation 1.11 is identical to regression 1.10, except that reported rates ($\hat{\tau}$) are

replaced with the respondent's actual rates (τ^*):

$$R_2 = \beta_0 + \beta_1 I_S + \beta_2 I_F + \beta_3 \tau_S^* + \beta_4 I_S \tau_S^* + \beta_4 I_F \tau_S^* + \beta_5 \tau_F^* + \beta_6 I_S \tau_F^* + \beta_7 I_F \tau_F^* \quad (1.11)$$

If respondents perfectly knew their taxes, and reported them without error, then regression 1.11 would be expected to yield the same results as 1.10. However, individuals may imprecisely know their tax rates, or their reports might be an imperfect function of the truth, so regression 1.11 will capture the amount the effect of the actual rates on overall reporting behavior, which might systematically differ from the effect of respondents' beliefs about taxes.

1.4 Results

1.4.1 Descriptive Results

Table 1.2 presents basic statistics about the actual tax rates faced by respondents and their errors in reporting tax rates for the five marginal-tax questions at the respondents' current income. The rates presented are for the five different sections: state, federal, total, round 1 and round 2. The error is equal to the reported tax rate minus the actual tax rate. Positive values represent a reported tax rate higher than the actual tax rate and negative values represent underestimates in reporting. The first row of errors contains the median error for each report. The median error in federal reports is to overestimate the federal income tax by 3 percentage points and, similarly, the median state response is overestimated by 2 percentage points. On the other hand, the total taxes paid by the household are consistently

underestimated: the median errors are 3 to 7 percentage points below the truth, depending on the section. The switch from over- to underestimating taxes when going from specific taxes to total taxes is suggestive that the median respondent may be ignoring payroll taxes in their estimate of total taxes; the payroll tax rate is approximately the size of the net difference in reports.

Table 1.2: Rates and errors in reported current tax rate, by type

Actual rates:	Federal	State		Total	
Median	15	4.6		28.7	
Mean	18.3	4.1		29.3	
25p	15	1.8		25.7	
75p	25	6		36.0	
Errors in reports:	**Federal**	**State**	**Combined**	**Round 1**	**Round 2**
Median	3.0	2.0	-3.4	-6.7	-4.3
Mean	6.4	11.9	-1.3	-4.4	-2.3
25p	-4.0	0.0	-11.7	-14.7	-12.7
75p	13.0	14.8	6.4	2.8	5.4
N	2,220	2,253	2,204	2,391	2,254

Notes: Error = Reported rate minus true rate. Positive value represent overestimates of true taxes, while negative values represent underestimates of true taxes.

1.4.2 Randomization check

In order to verify that randomization is successful and that my analysis sample selection did not lead to selection bias, I consider if round-1 answers vary by treatment assignment. Table 1.3 regresses the respondents' $\mathbf{R_1}$ answers on treatment assignment. Since treatment occurs after $\mathbf{R_1}$, it should not have an impact. Indeed, treatment does not have a statistically significant impact on the $\mathbf{R_1}$ answers.

Table 1.3: Randomization checks

VARIABLES	(1) R1
State treatment	-0.713
	(0.878)
Federal treatment	0.025
	(0.879)
Observations	2,326
R-squared	0.000

Notes: The dependent variable for the regression is the Round 1 total tax rate. Explanatory variables are the individual's treatment assignment. *** p<0.01, ** p<0.05, * p<0.1.

Another potential concern for the paper's analysis is that treatment changes the ordering of the specific rate questions. In the state treatment, the state rate is elicited before the federal rate, while in the federal treatment the federal rate is given before the state rate. To check if this ordering impacts the specific rates given, the reported state and federal rates were regressed on the treatment indicators in Table 1.4. As reported in the table, the treatment has no significant impact on the overall reported state rate. In fact, the treatment does not interact significantly with either of the reported rates that are used as inputs for further regressions.

Table 1.4: State and Federal reports, by treatment

VARIABLES	(1) State	(2) Federal
State treatment	1.614	-0.534
	(1.147)	(0.895)
Federal treatment	0.489	0.608
	(1.148)	(0.895)
Constant	15.087***	24.529***
	(0.809)	(0.631)
Observations	2,326	2,326
R-squared	0.001	0.001

Notes: The left column regresses the individual's reported state income tax rate on their treatment, while the right column repeats this analysis with the individual's reported federal rate as the dependent variable. *** p<0.01, ** p<0.05, * p<0.1.

1.4.3 Reduced-form Results

Table 1.5 contains the first evidence that there are different levels of salience between state and federal incomes taxes and that the experiment successfully manipulated salience. The table categorizes, by treatment group, individuals' self-reports of specific taxes included as part of their estimates of total taxes in round 2. Respondents are asked to provide this information immediately after round 2. As predicted, the state treatment made state taxes more salient: 72% of respondents in the state treatment include both federal and state taxes in their round 2 estimates, while only 35% of respondents do in the control. The federal treatment also seems to have made state taxes somewhat more salient (albeit, less salient than in the state group) and the proportion thinking about both types of taxes increases to 56%. These changes correspond with decreases in individuals who report only including federal taxes, shrinking from over half in the control to 17% and 35% in the state and federal treatments, respectively. If we exclude respondents in states

without income taxes, the share reporting both federal and state taxes increases by about 5 to 10 percentage points across treatments with corresponding decreases in those who report only federal taxes (as seen in Table 1.6).

Table 1.5: Taxes included in round 2 answers (reported)

	Treatment			
Thinking about:	State	Federal	Control	Total
Federal (N)	128	268	417	813
(% of column)	16.5	34.7	53.5	35.0
State	70	23	63	156
	9.0	3.0	8.1	6.8
Federal & state	556	463	269	1,288
	71.8	56.0	34.5	55.4
Neither	20	18	31	69
	2.6	2.3	4.0	3.0
Total	774	772	780	2,326

Notes: Table presents respondents' self-reports of which taxes they included in their answers in $\mathbf{R_2}$, categorized as to whether they include federal income taxes, state income taxes, both or neither (categories are mutually exclusive). The first three columns separate the sample by treatment condition, while the third presents the results for the entire sample.

Table 1.6: Taxes included in round 2 answers (reported), excluding zero tax states

	Treatment			
Thinking about:	State	Federal	Control	Total
Federal (N)	73	179	309	561
(% of column)	11.5	27.5	47.5	29.0
State	59	22	59	140
	9.3	3.4	9.1	7.2
Federal & state	484	434	258	1,176
	76.5	66.7	39.7	60.8
Neither	17	16	24	57
	2.7	2.5	3.7	3.0
Total	633	651	650	1,934

Notes: Replicates Table 1.5 but restricts sample to those in states with an income tax.

About 10% of respondents, across groups, answer in a manner inconsistent with my original expectation: either only including state taxes in their answer (which

runs counter to the hypothesis that federal taxes would be most salient) or including neither federal or state taxes in their round-2 reports (seemingly implying neither tax is salient to the individual). While some of these reports might be explained by an unexpected model of salience (or combination of salience parameters) for these individuals, it seems likely that many of these individuals remained confused (or inattentive) to the experimental task. Further evidence for confusion as an explanation is that the standard deviation in the errors in this group is substantially higher than in the other groups (36pp and 27pp, respectively).

The results from regression 1.9 are presented in column 1 of Table 1.7. The effect of being in the state treatment is, as expected, positive and significant: answers go up on average by 2 percentage points between round 1 and round 2 for participants in the state treatment relative to the control group. The impact of the federal treatment group is also positive: answers in the federal group went up 3 percentage points on average. These results provide further support that state taxes seem less salient to respondents. Reminding them that state taxes exist, either by having them think explicitly about state tax rates or by reminding them that federal tax rates are separate from total tax rates, increases their estimate of total taxes.[7]

Column 2 of Table 1.7 presents the results from regression 1.10. Both treatments increase the average reported rate, as the treatment effects are positive and significant; further indication they are now including state taxes. These treatment effects are once again roughly in line with average state income tax rates in the sample. I also find that the respondents' answers include a higher proportion of their reported federal tax rate than their reported state tax rate. The coefficient

[7]Adding additional individual controls for age, gender, education and income yield qualitatively identical, and numerically similar results for all regressions discussed in this section.

Table 1.7: Regression results

VARIABLES	(1) R2-R1	(2) R2	(3) R2
State treatment	2.392***	2.835**	0.507
	(0.807)	(1.152)	(2.252)
Federal treatment	2.745***	4.341***	5.903***
	(0.807)	(1.171)	(2.274)
State rate, reported		0.097***	
		(0.029)	
State Trt. * State rate		0.125***	
		(0.038)	
Fed. Trt. * State rate		0.047	
		(0.037)	
Federal rate, reported		0.680***	
		(0.035)	
State Trt. * Fed. rate		-0.127***	
		(0.047)	
Fed. Trt. * Fed. rate		-0.110**	
		(0.046)	
State rate, actual			0.821***
			(0.232)
State Trt. * State rate			0.383
			(0.314)
Fed Trt. * State rate			0.062
			(0.328)
Federal rate, actual			0.557***
			(0.079)
State Trt. * Federal rate			-0.006
			(0.105)
Fed Trt. * Federal rate			-0.174
			(0.111)
Constant	0.529	7.442***	11.866***
	(0.569)	(0.823)	(1.635)
Observations	2,326	2,326	2,326
R-squared	0.006	0.486	0.100

Notes: Each column represents a separate regression. The dependent variable for each regression is either the Round 2 total tax rate or the difference between Round 2 and Round 1. Explanatory variables are the individual's treatment assignment and either the individual's actual tax rates or the individual's reported tax rate (depending on the specific regression) interacted with treatment assignment. *** p<0.01, ** p<0.05, * p<0.1.

for state taxes is 0.10, meaning total tax answers only increase 0.1 percentage points for every 1 point increase in reported state rate. Federal taxes have a much higher coefficient of 0.7; federal tax beliefs have about seven times the impact of state tax beliefs. Furthermore, the state treatment more than doubles the degree to which the state tax is included in round-2 estimates: the interaction effect is 0.1. While the majority of the results point to state taxes being initially less salient, there is one piece of evidence that does not fit with the predicted model of salience: the impact of treatment on the effect of reported federal rates. Both treatment interactions are statistically significant and negative (-0.1). However, these negative interactions only slightly offset the main effect of treatment on the reported total rates which are 2.8 and 4.3 percentage points higher in the state and federal treatment groups, respectively. Overall, the treatment increased reports on net, which is consistent with the salience hypothesis, as is the fact that state reports influence the total less than federal reports.

Column 3 of Table 1.7 presents the results of regression 1.11. I find that the effect of actual tax rates are positive and significant. Respondents' round-2 answers reflect their true underlying tax rates to some degree, although less than 1. Interestingly, the coefficient is higher for state rates (0.8) than for federal rates (0.6). There is also no indication that the treatment changes the degree to which the actual tax rates are reflected in round-2 answers and none of the interactions are significant. The only significant estimate is the effect of being in the federal treatment which raises the average report by 6 percentage points. Note, the R^2 also drops to 0.1 from 0.49 in the prior regression. These results are consistent with the fact that individuals' state and federal rates contain a large amount of error, and therefore their reported rates may be a better source for testing the salience hypothesis.

Overall, regressions 1.9 and 1.10 provide the strongest reduced-form evidence of differential tax salience between federal and state taxes. The fact that reminding individuals of particular taxes changes their answers indicates that they were not initially considering all taxes. This evidence is further strengthened by the fact that reported state rates impact the respondents' overall report at roughly one-seventh the rate of reported federal taxes and that the impact of state rates over doubles when the respondent has been prompted to think about state rates specifically. These are also the two regressions which provide the cleanest test of tax salience as motivated by the theoretical model. Individuals' reports of total taxes should be most directly effected by their beliefs of the specific underlying tax rates. It is, in fact, the salience of these underlying beliefs that the experiment attempts to manipulate. Individuals' actual tax rates are a step removed from salience in the model and may be subject to biases due to individuals' errors, systematic or otherwise, in beliefs. Actual rates also suffer from an additional source of measurement error: they are correct for the "example" household, but due to misreported income or differences in tax-filing behavior may not be always correct for the respondent.

1.5 Model estimation

Next, I estimate a model of tax salience. This approach has the advantage of recovering direct salience-parameter estimates, whereas the reduced-form analysis only provided evidence for the relative effect of salience between the two taxes. The cost of this approach is that it requires much stronger, and potentially unrealistic, assumptions about how individuals perceive and report their tax rates. While the model matches the literature on tax salience and offers an appealing framework for

thinking about salience in this domain, it likely oversimplifies the unknown and heterogeneous nature of the true data-generating process by which individuals are reporting their tax rates. Therefore, estimates of the salience parameters should be taken only as further suggestive evidence of differences between the taxes, instead of as direct policy parameters.

To estimate the model, I begin by simplifying the error structure relative to the model in Section 1.3.1. Specifically, I assume that the individual error in perception of tax rate ($\delta_{L,i}$) is absorbed into the individual reporting error ($\epsilon_{L,i}$) when I elicit the tax rates. With this simplification, I can then estimate the two equations describing the individual's perceived federal and state tax rates directly:

$$\hat{\tau_{F_i}} = \beta_{F,0} + \beta_{F,1}\tau^*_{F_i} + \epsilon_{F_i} \tag{1.12}$$

$$\hat{\tau_{S_i}} = \beta_{S,0} + \beta_{S,1}\tau^*_{S_i} + \epsilon_{S_i} \tag{1.13}$$

Column 1 of Tables 1.8 and 1.9 contain the results from a regression of reported rate on actual rate for state and federal rates respectively. The state regression yields a coefficient that is approximately 1 in the linear specifications, while it is 0.5 for the federal rate (columns 2 and 3 present alternative quadratic and cubic specifications which yield similar predicted values). Fitted values from this regression are then used to directly estimate the salience parameters:

$$\tau_{T_i} = \beta_{T,0} + \theta^H_F\tau_{F_i} + \theta^H_S\tau_{S_i} + \epsilon_{T_i} \tag{1.14}$$

Table 1.10 presents the corresponding salience parameter estimates. I find qualitatively similar results in all three specifications (the primary linear specifica-

Table 1.8: Reported state rate as a function of actual state rate

VARIABLES	(1) State rate reported	(2) State rate reported	(3) State rate reported
State rate, actual	0.948***	1.297***	1.017**
	(0.160)	(0.264)	(0.513)
State rate squared, actual		-0.038*	0.010
		(0.023)	(0.079)
State rate cubed, actual			-0.001
			(0.002)
Constant	12.173***	11.706***	11.881***
	(0.798)	(0.846)	(0.890)
Observations	2,400	2,400	2,400
R-squared	0.014	0.016	0.016

Notes: Dependent variable is reported state tax rate (at current income). Each column presents the regression for a different specification, allowing the reported rate to be a linear, quadratic, or cubic function of actual rate, respectively. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.9: Reported federal rate as a function of actual federal rate

VARIABLES	(1) Federal rate reported	(2) Federal rate reported	(3) Federal rate reported
Federal rate, actual	0.512***	0.464***	0.540***
	(0.040)	(0.070)	(0.086)
Federal rate squared, actual		0.002	0.003
		(0.002)	(0.002)
Federal rate cubed, actual			-0.000
			(0.000)
Constant	15.271***	15.443***	14.492***
	(0.815)	(0.840)	(1.045)
Observations	2,367	2,367	2,367
R-squared	0.064	0.064	0.065

Notes: Dependent variable is reported federal tax rate (at current income). Each column presents the regression for a different specification, allowing the reported rate to be a linear, quadratic, or cubic function of actual rate, respectively. *** p<0.01, ** p<0.05, * p<0.1.

tion, as well as, the additional quadratic and cubic specifications). The state tax salience is lower than the federal tax salience with state tax salience being around 0.8, while federal tax salience is around 1.1. This finding means that individuals are overweighting federal taxes in their answers while underweighting state taxes. A surprising finding, given the reduced form results, is there is no statistically significant effect of treatment on any of the salience estimates; nothing is significant

Table 1.10: Model results, primary sample

VARIABLES	(1) R_2	(2) R_2	(3) R_2
θ_S	0.828***	0.790***	0.770***
	(0.232)	(0.203)	(0.216)
State trt.* θ_S	0.386	0.436	0.385
	(0.318)	(0.304)	(0.312)
Fed trt.* θ_S	0.062	0.065	0.064
	(0.332)	(0.307)	(0.294)
θ_F	1.108***	1.100***	1.076***
	(0.135)	(0.150)	(0.140)
State trt.* θ_F	-0.012	-0.075	-0.026
	(0.191)	(0.206)	(0.193)
Fed trt.* θ_F	-0.346*	-0.345*	-0.318
	(0.194)	(0.202)	(0.205)
State trt.	-3.845	-3.061	-3.452
	(5.777)	(5.796)	(5.558)
Fed trt.	10.480	10.417	9.765*
	(6.752)	(6.412)	(5.490)
Constant	-14.865***	-14.070***	-13.172***
	(4.287)	(4.165)	(3.957)
Observations	2,326	2,326	2,326
R-squared	0.100	0.100	0.101
Specification for reported rates	Linear	Quadratic	Cubic

Notes: Each column presents the structural estimate from a different model of salience. The left column presents the estimates when reported state and federal rates are a linear function of actual rates. The second column contains estimates when reported state and federal rates are a quadratic function of actual rates. Finally, column 3 allows reported rates to be a cubic function of actual rates. *** p<0.01, ** p<0.05, * p<0.1.

at the 5% level, and only two of the twelve total interactions are significant at the 10% level.

1.6 Subsample Analysis

As discussed previously, there is a large amount of error in individuals' reports of marginal tax rates, especially in the "tails." Through exploring the effect of this reporting error using simulation analyses, I found that the size of reporting error

(ϵ) in the reported rates could potentially confound measurements of salience.[8] My simulations show that large reporting error becomes a problem if (i) reported rates are correlated strongly with actual rates, and (ii) if reported rates increase as the underlying truth increases. The mistakes in reporting create a weaker relationship between the reported specific rates and the reported total rates which leads to smaller estimated coefficients for the reported state and federal rates. This concern motivates an additional set of analyses focused on samples with less error in their reports.

The preferred subsample, explored in this section, consists of individuals who made no bigger than a 20 percentage point error, in absolute terms, in reporting their state, federal or round-2 total rate (the rates used as components in the primary analyses). Additional subsamples explored, but not reported, are (i) those with less than 30 percentage point errors in the same three reports and (ii) those with correct answers to both training questions and (iii) error on their round-1 answers less than the median amount of combined error.[9]

Table 1.11 recreates and combines Tables 1.3 and 1.7 with the preferred subsample. A potential concern is that the subsample might suffer from some degree of selection bias since the treatment may influence the degree of error in round-2 and/or in the state and federal reports, and thus change who is removed from the sample. This concern is somewhat substantiated in column 1 as the treatment groups have lower round-1 answers than the control group. However this selection is in the opposite direction from the hypothesized treatment effect (which should then increase round-2 reports). After treatment the round-2 reports are significantly higher and the difference between round 2 and round 1 is higher than in

[8] For details, see Appendix A.1.

[9] Results in these subsamples generally qualitatively mirror the original sample or the preferred subsample.

Table 1.11: Randomization and results table, low error subsample

VARIABLES	(1) R1	(2) R2-R1	(3) R2	(4) R2
State treatment	-1.466*	2.682***	-0.374	0.198
	(0.843)	(0.678)	(1.022)	(1.233)
Federal treatment	-1.664**	3.354***	1.513	0.796
	(0.831)	(0.668)	(1.023)	(1.266)
State rate, reported			0.176***	
			(0.047)	
State Trt. * State rate			0.137**	
			(0.065)	
Fed. Trt. * State rate			0.079	
			(0.069)	
Federal rate, reported			0.857***	
			(0.033)	
State Trt. * Fed. rate			0.059	
			(0.047)	
Fed. Trt. * Fed. rate			-0.035	
			(0.045)	
State rate, actual				0.675***
				(0.131)
State Trt. * State rate				0.270
				(0.191)
Fed Trt. * State rate				0.011
				(0.186)
Federal rate, actual				0.635***
				(0.045)
State Trt. * Federal rate				0.014
				(0.062)
Fed Trt. * Federal rate				0.057
				(0.064)
Constant	23.366***	-0.450	4.739***	8.860***
	(0.594)	(0.477)	(0.733)	(0.901)
Observations	1,446	1,446	1,446	1,446
R-squared	0.003	0.019	0.633	0.416

Notes: Replicates and combines Tables 1.3 and 1.7 with the low-error subsample. *** p<0.01, ** p<0.05, * p<0.1.

the primary sample (and statistically significant). Despite these potential selection concerns, the potential bias from measurement error in the primary sample justifies an examination of the salience estimates in this low error group.

With these caveats in mind, column 5 provides additional evidence of salience effects. The reported state rate influences the round-2 report at a rate of 0.2, while the effect of federal rates is over four times as large (0.9). State treatment further

increases the salience of state rates, with the interaction term being 0.1, effectively doubling the contribution of reported state rates to the overall rate. Compared to the primary sample, there are also no interactions of the reported federal rate with the treatment. This reduction potentially reflects the reduced percentage of people that self-reported including only the state rate in round 2.

Table 1.12 estimates the structural model for this subsample. Again I find a lower salience for state taxes than for federal taxes (across all 3 specifications). In the subsample analysis the gap between the estimated θs increases and becomes more significantly different. The estimate for θ_S drops by about 0.1 and the standard errors around the estimates for the θs are cut approximately in half. Additionally, the model is able to much better explain the data as the R^2s increase from around 0.1 to 0.4. However, there remains no statistically significant effect of the treatment on the estimated salience parameters (once again contrary to the other descriptive evidence).

Table 1.12: Model results, low error sample

VARIABLES	(1) R_2	(2) R_2	(3) R_2
θ_S	0.680***	0.643***	0.633***
	(0.116)	(0.123)	(0.110)
State trt.* θ_S	0.289	0.287	0.296
	(0.190)	(0.188)	(0.190)
Fed trt.* θ_S	0.011	0.009	0.006
	(0.180)	(0.186)	(0.165)
θ_F	1.047***	1.063***	1.073***
	(0.069)	(0.066)	(0.069)
State trt.* θ_F	0.023	-0.003	-0.008
	(0.105)	(0.106)	(0.103)
Fed trt.* θ_F	0.094	0.093	0.073
	(0.106)	(0.102)	(0.093)
State trt.	-0.807	-0.398	-0.351
	(1.968)	(2.073)	(2.094)
Fed trt.	-0.088	-0.051	0.382
	(2.185)	(2.096)	(1.960)
Constant	-2.603*	-2.666**	-2.806*
	(1.440)	(1.332)	(1.496)
Observations	1,446	1,446	1,446
R-squared	0.416	0.416	0.418
Specification for reported rates	Linear	Quadratic	Cubic

Notes: Replicates Table 1.10 with low error sample. *** p<0.01, ** p<0.05, * p<0.1.

1.7 Heterogeneity in salience and tax knowledge

1.7.1 Heterogeneity in salience

In order to figure out the potential welfare implications of differential salience between taxes, it is important to understand the forces that might be driving salience. This section investigates several different potential hypotheses for factors that might drive salience. One hypothesis is that consumers are boundedly rational, so they pay attention to high taxes while ignoring smaller taxes. To examine this, I look to see to what extent high state or federal tax rates might affect tax reports and estimates of salience. It is also possible that more complicated tax sys-

tems lead to more consumer confusion and less salience as the consumer chooses to ignore the tax due to uncertainty in its true value. To investigate this hypothesis, I examine states with different degrees of complication in their tax schedules (ranging from a flat income tax to a fully progressive tax structure). Finally, it may only be more educated individuals or individuals with high incomes that bother to understand their income taxes, so I also examine salience by education and income level.

Table 1.13 explores some potential factors affecting salience. The first two rows report the reduced-form results for the effect of reported state and federal rates on the report in R_2 (equivalent to the same rows for regression 1.10). The third and fourth rows provide the structural estimates for θ_S and θ_F. Each column provides these estimates for a different subsample of the experiment. The columns contain the following samples: (1) individuals in states with the 50% highest tax rates; (2) individuals with at least a four-year college degree; (3) individuals in households with above median incomes; and (4) individuals that self-report having changed state or federal income brackets in the last 3 years.

In general, the smaller sample sizes lead to larger standard errors. While the individuals in higher-tax states seem to have generally higher impacts for state taxes (a higher reduced-form impact of reported state tax rates and model estimate for θ_S) large standard errors lead to a lack of statistically significant differences. Individual explanations, such as income and education, yield estimates that are not statistically significantly different than the full-sample estimates. Therefore, I am not able to identify an individual or policy-level factors that are clear determinants of tax salience.

As a robustness check, I also compare effects of tax salience across broad classes

Table 1.13: Comparing salience

	High State	High Education	High Income	Changed brackets	Full sample
Reduced form estimates:					
State rate, reported	0.158***	0.116***	0.099***	0.024	0.097***
	(0.04)	(0.040)	(0.037)	(0.052)	(0.029)
Federal rate, reported	0.605***	0.719***	0.771***	0.771***	0.680***
	(0.045)	(0.047)	(0.048)	(0.068)	(0.035)
Structural parameter estimates:					
θ_S	1.696***	1.057***	0.862***	0.538	0.828***
	(0.605)	(0.255)	(0.254)	(0.524)	(0.232)
θ_F	1.157***	1.233***	1.227***	1.762***	1.108***
	(0.228)	(0.172)	(0.387)	(0.295)	(0.135)
N	1,180	1,490	1,254	818	2,326

Notes: Top panel presents regression estimates for the effect of reported state and federal rates on round-2 rate, while bottom panel presents structural salience parameter estimates. Each column represents a different subsample that the estimation was performed on: (1) Individuals in states with the top half of tax rates; (2) individuals with more than a high school education; (3) Individuals with household income above median; (4) Individuals report having changed tax brackets in the last three years. *** p<0.01, ** p<0.05, * p<0.1.

of state-tax schedules. The results in column 1 of Table 1.7 are indicative of potential lower salience of state taxes, that are then included (and thus raise the rate) when made more salient by the treatment. To examine this effect more closely, Table 1.14 reruns the regression from equation 1.9, but splits the sample by the state tax structure. Column 1 is for the states that have a progressive tax schedule, where more income leads to higher taxes, and column 2 is for states where there is no state income tax. Making state tax more salient should not change answers, if respondents know the true rate is zero. As expected the treatment has no statistically significant effect on round-2 answers with no state tax, while it does have an effect in the progressive tax states.

Table 1.14: Treatment effects on change from R_1 to R_2, by state type

VARIABLES	(1) Prog. state tax	(2) No state tax
State treatment	2.564**	-1.718
	(1.266)	(2.060)
Federal treatment	3.880***	0.997
	(1.264)	(2.151)
Constant	26.551***	22.803***
	(0.908)	(1.500)
Observations	1,432	406
R-squared	0.007	0.004

Notes: Dependent variable: round-2 answer minus round-1 answer. Regresses change in rates on treatment assignment. Column 1 is the regression for individuals in progressive tax states while column 2 is for those in no tax states. *** p<0.01, ** p<0.05, * p<0.1.

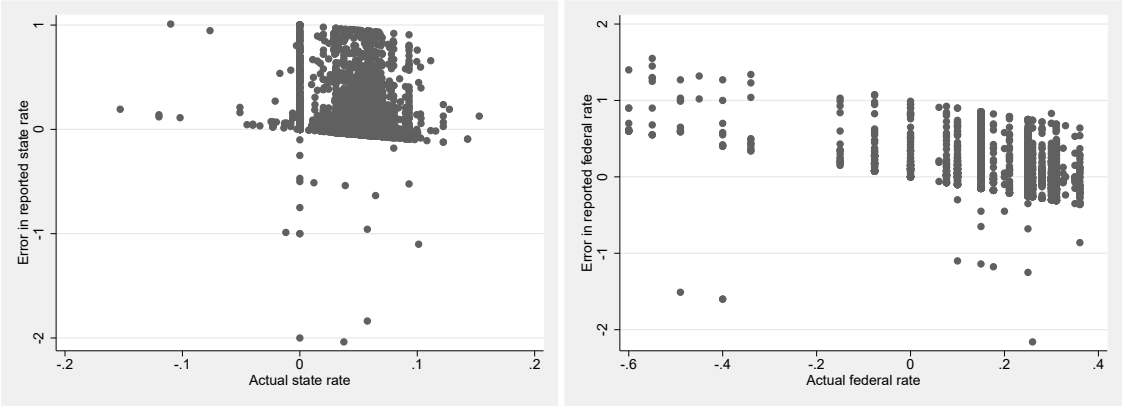
1.7.2 Heterogeneity in tax knowledge and errors

For an individual to react to a tax, the tax not only has to be salient to their decision-making, the individual also has to know the amount of the tax. The evidence is consistent that individuals generally know the sales tax they face, so salience seems to be a primary driving factor in the previously studied topic of individuals' responses to changes in sales versus excise taxes (Chetty et al., 2009; Taubinksy and Rees-Jones, 2017). However, the domain of income taxes is generally more complicated, and it is not clear that individuals completely understand the actual income taxes they face. This is shown in work by Chetty, Friedman and Saez (2013), who find that knowledge of features in the income tax schedule (and lack of knowledge) can drive earnings behavior. Some features of federal income tax knowledge have been studied, but this paper contributes new evidence on knowledge, its determinants, of state income taxes as well.

I begin by examining the overall nature of the error in tax knowledge. Figure 1.3 demonstrates that while both state and federal reported rates have high rates

of error, the magnitude of state errors are especially high relative to the true state rates. While many individuals have true federal rates in excess of 30%, the vast majority of state income tax rates are below 10% and, indeed, the median state tax rate is less than 5%. Therefore, despite having roughly equal levels of error, the ratio of actual rate to error is much lower for state rates than federal rates.

Figure 1.3: Actual rates and errors



Notes: Left panel presents comparison between actual state rates and errors in the reported state rates, while right panel does the same for federal rates.

Table 1.15 breaks down state errors further to see if observable features of the state tax code contribute to individuals' understanding of their state taxes. It might be the case that more complicated structures lead to more error and simple structures reduce errors in knowledge. To test this hypothesis, I separate individuals into one of three types of states: column 1 is states with progressive tax schedules, column 2 is states with no income tax, and column 3 are states with flat rates (a simpler tax schedule, where everyone pays the same income tax rate[10]). Those in progressive states, which have the most complicated tax schedules, have the highest rates of error by every metric. Their median error is 3 percentage points, while those in zero rate states have 0 median error and those

[10]This excludes the effects of some credits and deductions which may lead to slightly different marginal rates for limited income ranges.

in flat rate states only have 1 percentage point in error. Similarly, the mean error is 13 percentage points in progressive states, compared with 12 and 10 in zero tax and flat rate states, respectively. This evidence is consistent with increasing complications leading to increased errors.

Table 1.15: Error in state current tax rate, by state type

	Prog. states	Zero rate states	Flat rate states
Median	3.3	0	1.3
Mean	12.8	12.2	9.7
25p	0	0	-0.2
75p	16	11	12.8
N	1,427	409	564

Notes: Error = Reported rate minus true rate. Positive value represent overestimates of true taxes, while negative values represent underestimates of true taxes.

Table 1.16 looks at explaining the level of error for each type of report through observed individual characteristics (income, political affiliation, tax-preparation method, education) and tax-policy characteristics (state tax structure and average state rate). Each column has a different error as a dependent variable. Column 1 is the error in the round-1 report, column 2 is the round-2 error, column 3 is the total rate error, column 4 is the federal rate error, and column 5 is the state rate error. The regressions include a variety of individual and state characteristics to explain each round's error. Very few characteristics have any predictive power for explaining the errors in reports. Higher income is associated with lower errors across reports, except in the case of state reports where it actually increases the error in reports. For Round 1 and state taxes, having an professional prepare your tax is associated with a roughly 3-5 percentage point increase in error. For state reports, but not others, being in a flat-rate state is associated with a 3.2 percentage-point decrease in error. Low education (less than a high-school degree)

is associated with much higher error, but past high school the estimated effect of education on decreasing error is relatively flat.

A related issue in tax knowledge is that some people may misperceive their marginal tax rate as their average tax rate, misunderstanding the progressive nature of the income-tax schedule. Alternatively, the individual may just have better information on their average tax rates than their marginal tax rates and use it to proxy for their marginal rate (since average tax rate information is more readily apparent on pay stubs and tax filing documents). Table 1.17 provides some evidence that individuals do have a better sense of their average taxes than marginal taxes. This table shows the errors individuals make in reporting each kind of average rate. Generally, these errors are smaller than their marginal counterparts. Particularly, their total and state average rates which have median errors that are almost zero, and mean errors that are 2 to 3 percentage points. Additionally, for each type of rate, the 25-75 interquartile range (IQR) is smaller than its associated marginal IQR.

While the evidence suggests individuals are more accurate in identifying their average income-tax rates, it does not seem that individuals are fully misperceiving their marginal tax rate as their average tax rate. 58% of respondents report higher total marginal tax rates than average total tax rates. Similarly, 61% of respondents give answers consistent with progressive federal tax rates, and 55% of respondents do so for state tax rates. It appears the majority of respondents correctly perceive a difference between their average and marginal rates and directionally get this difference correct.

Table 1.16: Explaining error rates

VARIABLES	(1) R1	(2) R2	(3) Total	(4) Federal	(5) State
Income: 2nd quartile	-2.219**	-1.119	-2.103*	-1.477	4.767***
	(1.084)	(1.159)	(1.194)	(1.124)	(1.426)
3rd quartile	0.687	1.665	1.313	1.070	4.986***
	(1.024)	(1.095)	(1.128)	(1.062)	(1.347)
4th quartile	-0.044	1.005	0.095	-2.210**	3.635**
	(1.086)	(1.161)	(1.197)	(1.127)	(1.429)
Political affil: Democrat	-0.721	-1.495	-0.790	-1.502	-0.973
	(0.926)	(0.989)	(1.020)	(0.960)	(1.217)
Independent	-1.098	-1.138	-0.977	-1.303	-2.612**
	(0.980)	(1.048)	(1.080)	(1.017)	(1.289)
Libertarian	2.357	-2.322	-1.565	-0.223	-3.970
	(1.964)	(2.100)	(2.163)	(2.038)	(2.584)
Green	-1.594	5.352	11.924**	4.487	-7.453
	(4.301)	(4.598)	(4.738)	(4.462)	(5.658)
Tax prep: tax software	2.135	-0.174	0.774	-1.748	0.452
	(1.538)	(1.645)	(1.695)	(1.596)	(2.024)
Another HH member	1.476	-0.053	2.361	-2.188	1.075
	(1.962)	(2.097)	(2.161)	(2.035)	(2.581)
Professional	3.460**	2.750	2.848	1.870	5.223**
	(1.658)	(1.772)	(1.826)	(1.720)	(2.181)
Education: High school degree	-13.590	-15.032	-11.671	-11.839	-32.242***
	(8.557)	(9.148)	(9.425)	(8.877)	(11.256)
Some college	-14.024*	-15.903*	-9.660	-11.960	-34.373***
	(8.460)	(9.044)	(9.318)	(8.776)	(11.128)
Bachelor's degree	-14.957*	-15.629*	-10.666	-12.758	-37.006***
	(8.454)	(9.038)	(9.312)	(8.771)	(11.121)
Graduate or professional degree	-15.383*	-17.407*	-12.338	-14.856*	-40.233***
	(8.479)	(9.064)	(9.340)	(8.796)	(11.153)
Female	-0.183	-0.148	-0.452	0.603	1.607*
	(0.718)	(0.768)	(0.791)	(0.745)	(0.945)
Flat rate tax states	-0.400	-0.664	-0.973	-1.640*	-3.115**
	(0.947)	(1.012)	(1.043)	(0.982)	(1.245)
No tax states	-0.879	-0.999	-1.287	-2.100	-0.485
	(1.647)	(1.760)	(1.814)	(1.708)	(2.166)
Avg state tax: 2nd quartile	-0.574	0.227	-0.679	0.434	0.709
	(1.440)	(1.539)	(1.586)	(1.494)	(1.894)
3rd quartile	-1.953	-0.274	-1.862	1.233	-0.245
	(1.522)	(1.626)	(1.676)	(1.578)	(2.001)
4th quartile	-2.691*	-1.283	-3.332**	1.242	-0.019
	(1.499)	(1.602)	(1.651)	(1.555)	(1.972)
Constant	10.295	14.207	10.834	21.114**	44.302***
	(8.737)	(9.340)	(9.623)	(9.064)	(11.492)
Observations	2,297	2,297	2,297	2,297	2,297
R-squared	0.014	0.020	0.018	0.030	0.039

Notes: Each column regresses the error in reporting for a specific tax type (at current income) on individual characteristics and the characteristics of the state that they live in. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.17: Error in average current tax rate

	Avg. Total	Avg. Federal	Avg. State
Median	0.3	6.2	0.4
Mean	2.3	8.3	3.2
25p	-8.1	-0.3	-0.2
75p	8.7	13.7	3.8
N	2,290	2,291	2,291

Notes: Error = Reported average rate minus true average rate. Positive value represent overestimates of true taxes, while negative values represent underestimates of true taxes.

1.8 Conclusion

This paper's experiment and analysis finds evidence that suggests people react less to their state income taxes than federal income taxes. These findings are supported both in the reduced-form and structural-model results. These results are bolstered by the fact that the majority of individuals self-report only including federal rates in their overall answers if they have not been prompted to think about specific types of taxes, but this reverses to the majority including both federal and state income taxes when prompted to think about state income taxes first. While some of this effect could be driven by experimenter demand effects, the fact that thinking about federal taxes only (reminding them that different taxes exist, but not about state income specifically) creates about half of this reversal is evidence that this is likely not the only force at work.

These results are subject to some limitations. While the online sample provides a unique opportunity to experiment with a group that directly controls their marginal labor supply on a fine-tuned basis, the sample is not representative of the United States as a whole (neither demographically, nor with their labor-supply

decisions). While the reports in the experiment are incentivized, the experiment is unable to directly mirror the real-world setting of how individuals make labor-supply decisions. The large range of (and, in some cases, implausible or inconsistent) answers given in parts of the task indicate a potentially wide disparity in respondent understanding or effort towards the task. As discussed, this error could create some of the effects and patterns seen in the primary analysis. However, when a lower-error sample is used, the salience effects become even stronger so error in completing the task does not appear to be the prime driver of this paper's results.

These results have significant potential implications for optimal tax design as state income taxes seem to be potentially less distortionary than federal income taxes to behavior. Therefore, low-salience state taxes can be a useful lever for policymakers to manipulate. However, to fully understand the welfare effects of changes to state versus changes to federal income taxes, further exploration of the determinants of salience in this domain are needed. While this paper explored several potential drivers of salience, no strong evidence was found. This leaves the literature with two open questions: (1) which types of individuals find state taxes to be more or less salient and (2) what are the mechanisms underlying salience in this domain. The first question is important towards understanding the welfare effects of tax salience. The types of individuals find state taxes to be more or less salient changes the implication for optimal tax design. If the taxes are less salient for richer individuals (as in sales taxes), then raising state taxes might be less distortionary for high-income individuals; a potentially useful finding, given that high-income individuals are often found to have the highest earnings elasticity to tax changes. However, if the reverse is true, this would have the opposite implication for tax design.

The next question that needs to be explored is the mechanism underlying salience in this domain. Determining whether it is the relatively small size of state taxes, differences in media or political coverage, or some other factor that drives salience will be an important next step in order to determine the utility of differential salience as a policy lever. This paper finds some evidence that this effect comes, at least in part, from the complication in state tax schedules. However, other individual and structural components of salience need additional study.

CHAPTER 2

HOW DO PEOPLE SEE THEIR TAXES? EXPERIMENTAL EVIDENCE ON FEDERAL AND STATE INCOME-TAX PERCEPTIONS

2.1 Introduction

Individuals' perceptions of their taxes may influence their behavior. While it is often convenient to model perceptions as accurate, empirical evidence has increasingly shown that perceptions are often not free of systematic mistakes or biases (Feldman et al., 2016; Chetty et al., 2009). Individuals' incomplete perception of tax schedules has been shown to have substantial impact in actual consumer decisions, especially in areas such as the take-up of the earned income tax credit (EITC) where mistakes in perception may have large impacts on the household's total budget (Chetty et al., 2013). Determining individuals' understanding of the nature and general shape of the tax schedule is also important for analyzing their labor-supply decisions and our ability to measure the elasticity of taxable income using features of the tax schedule (Saez, 2010).

The literature has tried to characterize mistakes that individuals might make when perceiving their tax schedules. One area of particular focus has been understanding whether or not respondents correctly understand that there are differences between the marginal and average tax rates they face and, also, if they understand when each rate is applicable (Liebman and Zeckhauser, 2004). The evidence on this topic is mixed: see Taubinksy and Rees-Jones (2017) for an example where participants appear to be "ironing" (using the average rate in place of the marginal rate) and Gideon (2017) for an example of participants correctly

perceiving the difference between average and marginal rates. A related, but less studied, area of active research is trying to measure individuals' perceptions of the broader tax schedule, and not just the rate individuals themselves face (Taubinksy and Rees-Jones, 2017). This research investigates whether individuals understand the existence of kinks in the tax schedule and the broader, progressive nature of the federal tax system in the United States. Up until now, the federal tax schedule has been the primary focus of research in the literature on tax perceptions. However, state income-tax schedules provide a rich diversity of income-tax regimes and rates. The heterogeneous nature of state tax schedules and their effects on individuals' perceptions has not been directly addressed in the literature.

The first aim of the paper is to determine whether individuals correctly distinguish between marginal and average tax rates and to compare the degree to which this understanding differs between state and federal taxes. The second aim is to analyze whether individuals understand the progressive nature of the federal income tax schedule and then compare the magnitude of their perceptions of progressivity with the actual progressivity in the tax schedule; and, to do likewise for states with progressive tax schedules. The final aim of the paper is to build and empirically estimate a model of federal tax perception designed to asses how people's perceptions of their average and marginal rates depend on their true average rates and marginal rates.

This paper analyzes data collected using Amazon's MTurk platform from over 3,000 respondents drawn from the United States. These individuals were incentivized to answer questions about their perceived marginal and average tax rates at a variety of income levels and for three types of taxes: federal, state, and total (inclusive of state, federal, and payroll). While the primary design of the experi-

ment was to measure and manipulate tax salience (as analyzed in Chapter 1), in this chapter we use its questions on tax knowledge to analyze perceptions of state and federal tax schedules.

We begin by exploring whether people recognize two basic patterns of progressive tax schedules: (1) in virtually all instances average tax rates are less than marginal tax rates, and (2) in many instances, marginal tax rates are higher at higher incomes. We find the bulk of the evidence to be consistent with most individuals having correct perceptions of the progressive nature of state and federal income tax rates (where applicable). For both state and federal income taxes, around 62% of individuals report higher marginal than average rates (consistent with a progressive tax scheme). Additionally, looking within subjects at both types of taxes, individuals tend to report higher taxes when asked about tax rates at higher incomes and lower taxes at lower incomes: around 30-40%, depending on the type of tax and distance away from current income, report changes consistent with a progressive tax schedule. In fact, the evidence suggests that individuals overestimate the progressive nature of both state and federal tax schedules, reporting that the tax rate changes much more quickly than the actual rates change, relative to changes in income. Between three and seven times as many people report a change in the rate, compared with those who actually experience a change in tax rate. Across all of these metrics, state and federal perceptions seem broadly similar.

The paper then develops a model of tax perceptions. In this model, an individual has perceptions of their marginal and average tax rates. These perceptions may depend on either their true marginal or average tax rates (or both). The model will then be estimated using mixture techniques to determine the types of

respondents in the data (e.g. do some respondents only use their true marginal rate as the basis of both their marginal and average answers). This model allows us to see if individuals seem to be confusing average and marginal tax rates or correctly perceiving the difference between the two. Here the results indicate that individuals appear to be neither "ironing" (always using their average rate) nor "spotlighting" (always using their marginal rate) in ways that behavioral models have predicted; nor are they correct as assumed by traditional models (fully separating their estimates of average and marginal rates). Instead, we find evidence consistent with both average and marginal rates influencing both types of answers.

Finally, the paper focuses on two groups of individuals in the federal income-tax brackets containing the most respondents in the data: those that pay 15% or 25% in marginal federal income taxes. Focusing on these two groups, which contain about half the total sample, allows us to eliminate noise due to individuals in brackets where we do not observe many people or due to small notches in the federal income-tax schedule (caused by the phase-in or phase-out of deductions and/or credits, which could increase the effect of income measurement error on our estimates). Within these two groups we can then examine how reporting behavior changes with distance from three changes in the tax schedule (the start of the 15% bracket, the transition from 15% to 25%, and the end of the 25% bracket). These reports exhibit a phenomenon similar to ironing (but at the population level instead of individual level): the marginal rate reported grows in a linear fashion (akin to how the underlying average federal tax rate is increasing), instead of in discontinuous jumps as the true marginal rate does. Reports go from over-reporting their true tax rate among the 15% group, on average, to under-reporting in the 25% group, on average. Estimating the earlier model of perception among this group leads to broadly similar results: there appears to be no type of individual that purely

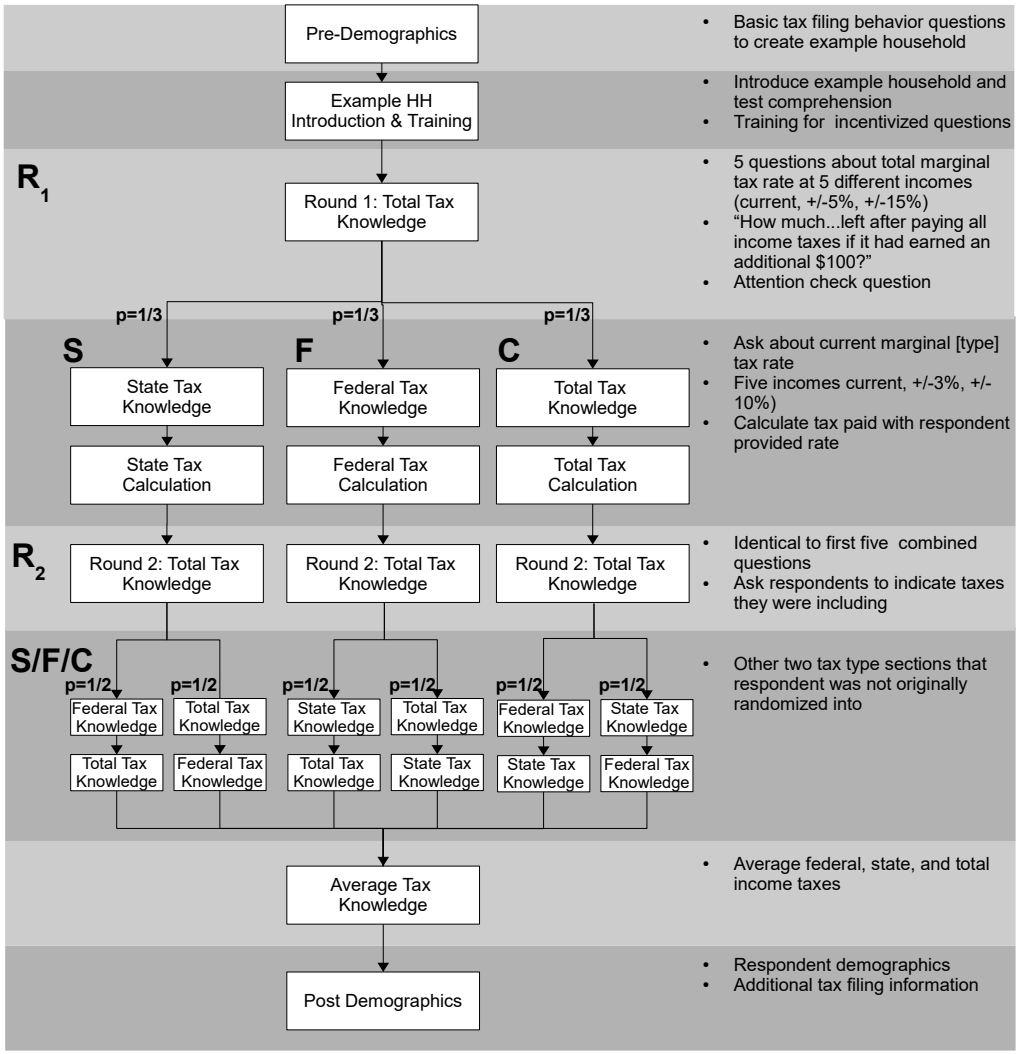
"irons," purely "spotlights," or is fully correct.

The evidence presented in this paper suggests several implications for individual behavior. First, given the broad similarities in reporting behavior between state and federal tax rates, there do not seem to be dramatically different perceptions of state and federal tax schedules. Respondents seem to view, and therefore potentially react to, the taxes similarly. This equivalence is potentially useful for state policymakers, who may wish to use elasticity estimates from federal income taxes when considering policy. Additionally, the mix of evidence, with findings for both progressivity and "ironing," suggests that traditional models of tax perceptions may be too restrictive. Current theoretical models of perception do not explain the data well. The gap between individuals' perceptions of taxes and the actual tax changes may also be a potential explanation for the empirically observed lack of bunching at kinks in the tax schedule, which runs contrary to previous theoretical predictions.

2.2 Experiment

We designed an experiment to measure individuals' perceived marginal tax rates and to manipulate the salience between state and federal income taxes. Figure 2.1 illustrates the experiment's overall design. This design and the experiment's sample was discussed in full detail in Chapter 1, Section 1.2. The experiment's core is divided into five primary sections, each one containing five questions. The questions ask what the respondent thinks the marginal tax rate is at a given level of income for a household designed to be similar to their own in tax filing behavior. The primary question gave the respondent an initial income level and asked "How

Figure 2.1: Experiment design



Notes: This figure illustrates the experimental design. Here we provide All participants begin with the same three sections, before being randomized to one of two treatment groups (State or Federal) or the control group. By the end of the experiment, all participants receive each of the three main sections $\mathbf{R_1}$, **S**, **F**, **C**, and $\mathbf{R_2}$. However, randomization controls the order that they receive **S**, **F**, and **C**.

much more would the [example household] have, after paying **all income taxes** on the additional earnings, if it had made an additional $100?", where the specific type of tax changes by section. (Refer to Appendix B for the full survey text.)

This paper focuses on measurements made in three of those sections: labeled **F** (Federal), State (**S**), and Combined (**C**) in the figure. These sections elicit respondents' beliefs about their marginal federal, state, and total income tax rates, respectively. Each of these sections elicit beliefs about these rates at five different income levels, centered at the respondents' reported household income level. The other four incomes where the respondents' beliefs are measured are at current income plus 3%, plus 10%, minus 3%, and minus 10%. The other questions of interest to this paper are the average tax-rate measurements, which ask about average state, federal, and total tax rates. This measurement occurs after all the marginal rates are measured (labeled section **A**). Average tax rates are only measured for the respondents' current household income level, not at the other four income levels. Summary statistics for each of the marginal and average reports are presented in Table 2.1, using the analysis sample. This sample is chosen for their relatively low error in reporting their current federal marginal tax rate, as will be described in more detail in Section 2.4.2.

Table 2.1: Reported state and federal tax rates, by income level

	Mean	Median	SD
Reported Marginal Federal Tax			
1.10I	20.3	20.0	16.9
1.03I	19.1	19.0	16.1
1.00I	17.2	16.0	17.2
0.97I	17.9	16.0	16.3
0.90I	17.2	15.0	16.7
Reported Marginal State Tax			
1.10I	12.7	6.0	21.1
1.03I	12.3	6.0	19.8
1.00I	11.7	6.0	19.8
0.97I	12.1	6.0	21.3
0.90I	11.7	5.0	19.7
Reported Marginal Total Tax			
1.10I	25.1	25.0	19.4
1.03I	24.9	23.6	18.9
1.00I	23.0	23.0	18.0
0.97I	22.9	22.0	18.3
0.90I	21.5	20.0	18.4
Reported Average Federal Tax			
1.00I	13.6	12.1	12.6
Reported Average State Tax			
1.00I	5.3	3.1	9.9
Reported Average Total Tax			
1.00I	17.5	16.0	16.7

Notes: N=2,326. Provides summary statistics for reported state, federal, and total marginal tax rates at each of the five income levels ranging from 110% of income (1.10I) to 90% of income (0.9I). Additionally, reported average rates at current income level (1.00I).

2.3 Results

2.3.1 Relationship between ATR and MTR

A key feature of income taxes is that the marginal tax rate faced by an individual is virtually always higher than the average tax rate faced by that individual.[1] Even for individuals in the lowest income-tax bracket or those who live in states with only one marginal tax rate, the effect of deductions and exemptions make an individual's average rate lower than that individual's marginal rate. We first test

[1]The only notable exceptions to this pattern are individuals in states with no income taxes and for individuals in small income notches with low marginal rates due to the phase-in of a credit or deduction.

whether individuals are aware of this feature of tax schedules.

Table 2.2 presents the general patterns found in individuals' answers for their marginal tax rate (MTR) and average tax rate (ATR) for their state, federal, and total tax answers. Here we find broadly consistent patterns across the three tax rates, with the majority of individuals correctly reporting marginal tax rates that are higher than their reports for average tax rates. For both federal and total income tax rates, around 55% of individuals report higher marginal tax rates than average tax rates, and only around 20-25% of respondents report lower marginal tax rates than average tax rates.

Table 2.2: Comparing average and marginal rate answers

	State	Federal	Total	No-tax states	Progressive states
$MTR > ATR$	49.1	53.6	54.1	25.4	58.2
$MTR = ATR$	34.0	27.0	21.0	67.4	23.1
$MTR < ATR$	16.9	19.4	24.9	7.1	18.7
N	2,326	2,326	2,326	393	1,235

Notes: Compares individuals' marginal tax rate answer (MTR) and average tax rate answer (ATR) for three categories of income taxes: state, federal, and total tax. All answers are at the respondent's current income level. Each cell tells the proportion of the sample whose answers exhibited the pattern given by the column label. Zero-rate states are those with no income tax for all individuals. Progressive-rate states are those that have increasing income taxes for higher incomes.

When looking at the overall pattern in state reports, only 49% of respondents report marginal rates in excess of their reported average rates. However, states can have one of three categories of income tax schedules: progressive (higher incomes lead to higher tax rates), flat (the same marginal rate across all incomes), or no state income tax. If we examine only individuals in progressive income tax states, then the pattern of reports more closely matches federal reports: around 60% of respondents report higher MTR than ATR, and only 19% of respondents report an MTR less than ATR. Comparing this with respondents in states with zero income taxes, where 67% of respondents correctly report an MTR equal to ATR.[2]

[2]In states with flat income tax schedules, there was a similar pattern to the progressive

2.3.2 Perceptions of progressivity

A second key feature of the federal tax system and of many states is that taxes are progressive in the sense that higher incomes have a higher MTR. We next examine whether individuals are (i) aware of this feature, and (ii) aware of its magnitude. The experimental design also allowed us to measure individuals' beliefs about the tax schedule near their current household income. We elicited individuals' beliefs about the marginal state and federal tax rates at each of five different income levels: their current income, current income plus 3%, current income plus 10%, and the equivalent amounts below their current income. Using these measures we could see respondents' perceptions of the state and federal income tax schedules near their income and compare it with the true tax schedules.

Table 2.3 compares the directional relationship between the MTR reported at the household's current income level with the MTR reported at the other four income levels (two below current income and two above). In the top panel, each column compares the MTR at the given income level with the MTR reported at the household's current income level. For instance, the top left cell reports the proportion of respondents who gave a state MTR rate at the income minus 10% level that was lower than the rate they reported at their household's current income level, which was 27% of respondents. In the cell below, we find that 64% of people reported the same state MTR at both their current income level and the income minus 10% level. Moving in the other direction, we see that at income plus 10% about 28% of people reported a state MTR that was higher than the one they reported at their actual income rate, while 64% of respondents reported the same state MTR at both income levels. The rightmost four columns repeat the

states 30% of respondents reporting equal marginal and average tax rates and 50% reporting $MTR > ATR$ for state rates.

comparisons using the individuals' federal MTR reports. Finally, the lower panel reports the true relationships between the actual MTRs at each income level. For instance, in the bottom left cell we see that for 12% of respondents, the state tax rate was actually lower at current income minus 10% than it was at the household's current income level, while the fourth column shows that 10% of households would have had higher marginals at their current income plus 10% than at their current income.

Table 2.3: Perceptions of local tax schedule

State reports				Federal reports			
$\tau_{D10} < \tau_C$	$\tau_{D3} < \tau_C$	$\tau_{U3} > \tau_C$	$\tau_{U10} > \tau_C$	$\tau_{D10} < \tau_C$	$\tau_{D3} < \tau_C$	$\tau_{U3} > \tau_C$	$\tau_{U10} > \tau_C$
27.1	17.8	18.4	27.7	35.0	22.1	25.8	39.5
$\tau_{D10} = \tau_C$	$\tau_{D3} = \tau_C$	$\tau_{U3} = \tau_C$	$\tau_{U10} = \tau_C$	$\tau_{D10} = \tau_C$	$\tau_{D3} = \tau_C$	$\tau_{U3} = \tau_C$	$\tau_{U10} = \tau_C$
64.0	69.0	70.6	64.0	54.6	64.4	64.5	53.4
$\tau_{D10} > \tau_C$	$\tau_{D3} > \tau_C$	$\tau_{U3} < \tau_C$	$\tau_{U10} < \tau_C$	$\tau_{D10} > \tau_C$	$\tau_{D3} > \tau_C$	$\tau_{U3} < \tau_C$	$\tau_{U10} < \tau_C$
8.9	13.2	11.1	8.3	10.3	13.5	9.7	7.2
Actual rate				Actual rate			
$\tau^*_{D10} < \tau^*_C$	$\tau^*_{D3} < \tau^*_C$	$\tau^*_{U3} > \tau^*_C$	$\tau^*_{U10} > \tau^*_C$	$\tau^*_{D10} < \tau^*_C$	$\tau_{D3} < \tau^*_C$	$\tau^*_{U3} > \tau^*_C$	$\tau^*_{U10} > \tau^*_C$
11.6	4.9	3.5	9.5	17.0	6.0	4.1	15.2

Notes: $N = 2,326$. The top panel compares individuals' marginal tax rate reports at 5 different income levels: current income (τ_C), down 10% (τ_{D10}), down 3% (τ_{D3}), up 3% (τ_{U3}), and up 10% (τ_{U10}). The bottom panel compares the actual tax rates (τ^*) at the same 5 income levels.

Comparing the patterns across state and federal reports, we find broadly similar results. Overall, respondents seem to have a sense of the progressive nature of most income-tax schedules. In state reports, around 27% of people give a lower state MTR at the lowest income they see, while 28% report a higher state MTR at the highest income they see. Only about 10% of respondents report the opposite pattern of higher state-tax rates at lower incomes or lower-state tax rates at higher incomes. In the federal reports, we observe a similar pattern with 35% of respondents reporting lower federal MTRs at the lowest income level and 40% reporting higher federal MTRs at the highest income. Likewise, only about 10%

respondents give answers inconsistent with a progressive federal-tax scheme.

Another piece of evidence that individuals have a sense of the progressive nature of federal (and many states') tax schedules is that for intermediate values of income (plus or minus 3% from current household income) an intermediate number of respondents report rates that are higher or lower, respectively, than the rate reported at their current income level. When we limit the sample to individuals in states with progressive taxes in Table 2.4, we find even more similar patterns, where state behavior looks similar to federal behavior, although still with generally slightly lower proportions.

Table 2.4: Perceptions of local tax schedule, progressive states

State reports				Federal reports			
$\tau_{D10} < \tau_C$	$\tau_{D3} < \tau_C$	$\tau_{U3} > \tau_C$	$\tau_{U10} > \tau_C$	$\tau_{D10} < \tau_C$	$\tau_{D3} < \tau_C$	$\tau_{U3} > \tau_C$	$\tau_{U10} > \tau_C$
33.1	20.9	22.8	33.8	34.9	22.1	25.8	39.5
Actual rate				Actual rate			
$\tau^*_{D10} < \tau^*_C$	$\tau^*_{D3} < \tau^*_C$	$\tau^*_{U3} > \tau^*_C$	$\tau^*_{U10} > \tau^*_C$	$\tau^*_{D10} < \tau^*_C$	$\tau_{D3} < \tau^*_C$	$\tau^*_{U3} > \tau^*_C$	$\tau^*_{U10} > \tau^*_C$
20.9	8.8	6.0	15.8	17.0	6.0	4.1	15.2

Notes: $N = 1,235$ for state reports and $N = 2,326$ for federal reports. Recreates the state portion of Table 2.3 but limits the sample to states with progressive tax schedules.

While individuals, in the aggregate, seem to understand the progressive nature of the tax structure, they also seem to overestimate the nearness of the proximity to the next kink. 27-35% of individuals report a lower MTR when their income is lowered by 10%, depending on the rate type, but only 12-17% of respondents would actually have the corresponding rate change. Similarly, respondents overestimate the quickness at which the tax rate changes when their income increases by 10%: 28-40% of respondents report increases, with only 10-15% of respondents actually experiencing tax-rate increases. Of the individuals who change their rates as income moves up or down, around 35% (for state) and 40% (for federal) make changes in both directions (i.e. report a lower rate at a lower income and a higher

rate at a higher income), while the other 60-65% change their tax rate answer in only one direction.

2.3.3 The proximity of kinks

Another interesting question is the relationship between the proximity of kinks and tax perceptions. That is, does being near a kink change how individuals view their tax rates? It is possible that individuals who are near a kink are either more or less aware of the marginal tax rate they face. They might be more aware if the effect of crossing a kink and facing a higher rate has made it especially salient to them, or the rate might be more salient to individuals who are aware of a nearby kink and are trying to stay in the lower marginal-tax bracket (which might give them a heightened understanding of their current rate). On the other hand, it might be lower if they have recently moved to a higher bracket and they have not yet fully incorporated the higher rate into their perceptions.

To investigate potential effects of being near kinks on perceptions, we estimate a straightforward regression of perceptions on actual rate separately for those who are near or far from a kink. That is we regress:

$$\hat{\tau_F} = \beta_0 + \beta_1 \tau_F + \epsilon$$

separately for those who are either "near" or "far" from a kink. We try two different measures of distance: a threshold of 10% of income away from a kink or a threshold of 20% of income away from a kink.

Table 2.5 presents the results from these regressions. Broadly speaking, we find

similar patterns of results for both those near or far from kinks, regardless of the cutoff used. In each case we cannot reject that the coefficients estimated for each sample are the same. Thus we conclude that proximity to kinks does not appear to be directly related to perceptions in the full sample. However, as we will show in Section 2.4.1, this may be a power limitation when trying to examine every kink in the sample, instead of focusing on kinks with sufficient numbers of individuals nearby.

Table 2.5: Perceptions relative to kinks

Near kink (10%)		Far from kink (10%)	
τ_F^*	0.253***	τ_F^*	0.236***
	(0.032)		(0.025)
Constant	12.76***	Constant	12.95***
	(0.708)		(0.559)
N	875	N	1,520
Near kink (20%)		**Far from kink (20%)**	
τ_F^*	0.202***	τ_F^*	0.209***
	(0.022)		(0.026)
Constant	13.80***	Constant	14.01***
	(0.460)		(0.520)
N	1,451	N	806

Notes: Results from four different regressions, each on a different subsample, of federal tax rate answer on actual federal tax rate. Each regression is based on distance from closest kink in the federal tax schedule with the top two based on a distance of 10% of income from the nearest kink and the bottom two based on a distance of 20% of income from nearest kink.

2.4 Estimating a model of perception

The prior literature has suggested a number of ways that people might perceive their taxes. One is that individuals might correctly perceive taxes (knowing the difference between the average and marginal tax rates that they face). The second is that they might "iron" and use the average tax rate in place of their marginal

tax rate. The third is that they might "spotlight" and use their marginal tax rate in place of their average tax rate. In this section we develop a model that nests each of these three types.

2.4.1 Model description

We elicit two federal tax rates of interest for each respondent: their current marginal federal tax rates and their current average federal tax rates. In standard economic analyses, each of these reports would typically be assumed to be reports of the respondent's actual marginal and average rates, with potentially some noise. However, as suggested by the literature, respondents might not always distinguish between marginal and average tax rates. Instead, they may report both their average and marginal rates as a function of either one or the other. We use a flexible model to estimate individuals' reports. Their reports may be a function of both their actual average and marginal tax rates:

$$\hat{\tau^M} = \beta_{0,M} + \beta_{1,M}\tau^M + \beta_{2,M}\tau^A + \epsilon_1 \tag{2.1}$$

$$\hat{\tau^A} = \beta_{0,A} + \beta_{1,A}\tau^M + \beta_{2,A}\tau^A + \epsilon_2 \tag{2.2}$$

Here $\hat{\tau^M}$ and $\hat{\tau^A}$ are the respondents' reports for marginal and average tax rates, τ^M is the individuals' actual marginal tax rate, and τ^A is the individuals' actual average tax rate .

The literature suggests different individuals may report rates in systematically

different ways (Liebman and Zeckhauser, 2004; Rees-Jones and Taubinksy, 2016). In particular, there are three common suggestions for types of individuals. Type 1 incorrectly base both their reports of average and marginal rate on their average rate (such that $\beta_{1,A}$ and $\beta_{1,M}$ are equal to zero):

$$\hat{\tau^M} = \beta_{0,M} + \beta_{2,M}\tau^A + \epsilon_1$$

$$\hat{\tau^A} = \beta_{0,A} + \beta_{2,A}\tau^A + \epsilon_2$$

Type 2 reports both as a function of their actual marginal tax rate (such that $\beta_{2,m}$ and $\beta_{2,A}$ are equal to zero):

$$\hat{\tau^M} = \beta_{0,M} + \beta_{1,M}\tau^M + \epsilon_1$$

$$\hat{\tau^A} = \beta_{0,A} + \beta_{1,A}\tau^M + \epsilon_2$$

And finally, type 3 reports using the correct rate as the source of each report ($\beta_{2,M} = \beta_{1,A} = 0$):

$$\hat{\tau^M} = \beta_{0,M} + \beta_{1,M}\tau^M + \epsilon_1$$

$$\hat{\tau^A} = \beta_{0,A} + \beta_{2,A}\tau^A + \epsilon_2$$

Note that all three of these cases are embedded in the initial model.

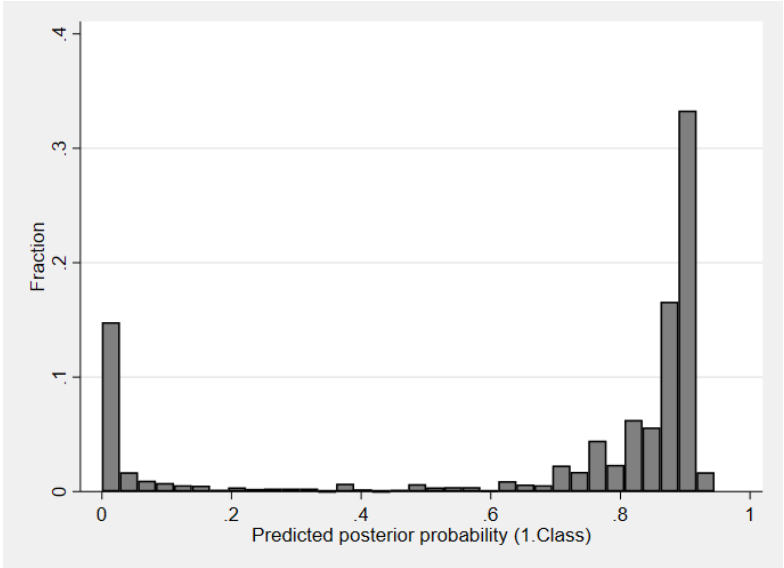
2.4.2 Mixture results

We estimate the model of individuals' marginal and average tax perception (as described by equations 2.1 and 2.2) using Stata 15's finite mixture models. As discussed in Section 1.2 this sample has a high number of individuals reporting clearly erroneous tax rates either due to confusion about the experiment or about the tax code. In Chapter 1, we limited the sample using cutoffs in reporting error. This chapter adopts a different approach and estimates a first-stage mixture model based on individuals' reports of federal rates.[3] We estimate a simple mixture model of federal reports as a function of actual marginal and average federal rates, allowing for two types of respondents. One type clusters around individuals who reported federal rates that were in the realm of actual federal rates, while the other contained individuals with extremely high reports for federal rates. Figure 2.2 presents the calculated posterior probabilities of each individual's membership in each group (type A are individuals reporting "reasonable" rates while type B are individuals reporting "high" rates). The 2,326 individuals that were most likely type A (i.e. had probabilities of being type A in excess of 50%) and who completed the average tax section were kept in the primary analysis sample.

The model from Section 2.3.3 is then estimated using a finite mixture model. Inputs are the individuals' reported marginal and average federal tax rates and their corresponding actual rates (calculated using Taxsim 9 as discussed in Section 1.2). The estimation allows for three types of respondents, but did not constrain the coefficients or their relationships (i.e. the hypothesized types were not imposed as part of the model estimation). We allow for three types in the estimation in order to attempt to discover the three types typically discussed in the literature (as

[3]Cuts using state or total rates instead of federal rates yield similar qualitative results. Additionally, similar results are found when imposing cutoffs on errors in reporting.

Figure 2.2: Type probabilities



Notes: Individual posterior probabilities of type for initial sorting model. Those with low probabilities of being type A are those reporting very high federal marginal rates (e.g. near 100%).

discussed in the previous section). Table 2.6 presents the three sets of coefficient estimates for each of the three types estimated in the sample. Each panel contains the results for one type of individual and the six coefficient estimates associated with that type. The first set of coefficients in each panel shows the relationship between an individual's reported marginal federal tax rate and their underlying actual marginal rate and actual average rate (i.e. $\beta_{0,M}$, $\beta_{1,M}$ and $\beta_{2,M}$), while the second set shows the relationship between reported average federal tax rate and the underlying actual rates (i.e. $\beta_{0,A}$, $\beta_{1,A}$ and $\beta_{2,A}$). Finally Table 2.7 presents the overall marginal posterior probabilities for membership in each type.[4]

None of the three types directly correspond to the types hypothesized by the literature. However, type 1 seems to be the closest. For the type-1 individual, the coefficients on the actual marginal rates are insignificant (corresponding with

[4]We tried to estimate a similar model using the state reports, but we were unable to get the model to successfully converge for three types.

Table 2.6: Mixture model results

Type 1:					
Dep var: $\hat{\tau_F^M}$			Dep var: $\hat{\tau_F^A}$		
	τ_F^M	-0.01		τ_F^M	-0.00
		(0.032)			(0.035)
	τ_F^A	0.134***		τ_F^A	0.175***
		(0.031)			(0.034)
	Constant	6.78***		Constant	6.11***
		(0.678)			(0.693)
Type 2:					
Dep var: $\hat{\tau_F^M}$			Dep var: $\hat{\tau_F^A}$		
	τ_F^M	0.168***		τ_F^M	0.074***
		(0.024)			(0.026)
	τ_F^A	0.149***		τ_F^A	0.169***
		(0.023)			(0.025)
	Constant	18.66***		Constant	13.3***
		(0.459)			(0.482)
Type 3:					
Dep var: $\hat{\tau_F^M}$			Dep var: $\hat{\tau_F^A}$		
	τ_F^M	-0.067		τ_F^M	0.270
		(0.115)			(0.154)
	τ_F^A	0.249*		τ_F^A	0.012
		(0.134)			(0.200)
	Constant	16.2***		Constant	80.56***
		(1.91)			(2.36)
		$N = 2,326$			

Notes: Reports the mixture model results for each of the three types estimated, separated by the three horizontal panels. The left columns presents the effect of actual marginal (τ_F^M) and actual average (τ_F^A) federal tax rates on the respondents' reported marginal tax rate ($\hat{\tau_F^M}$), while the right column does the same for the respondents' reports of average federal tax rate ($\hat{\tau_F^A}$).

the predicted type of $\beta_{1,M} = \beta_{1,A} = 0$), while the coefficients on the average tax rate are positive and significant. However in both cases, the estimated coefficients are significantly less than the estimate of 1 which would be expected if individuals were fully incorporating their average rates into their estimates. We also note the relatively low estimates for β_0 for both dependent variables, combined with the relatively small effect of both actual tax rates on the report, indicates individuals

Table 2.7: Latent class probabilities

Class	
1	0.337
	(0.021)
2	0.650
	(0.021)
3	0.013
	(0.002)

Notes: Posterior probabilities of class membership from mixture model estimated in Table 2.6.

of this type have relatively low reported marginal and average tax rates ($< 15\%$). As seen in Table 2.7, the probability of being classified as this type in the sample is about 34%.

Type 2 does not correspond with the potentially expected individual using their marginal tax rate for all the answers, nor a fully rational individual using their actual marginal rate to generate marginal reports and their actual average rate to generate their average reports. Instead, for this type, individuals' reports for both rate types are a function of both actual rates. All four coefficient estimates are positive and significant, albeit low, in the range of 0.07–0.17. The estimates for the constant are significantly higher than for a type-1 individual (13.3 and 18.7, respectively). Combined with the small but positive effects of actual rates, most individuals in this group have federal tax reports in the moderate range (15-25%). This group had the substantially highest probability of being classified, about 65%.

Finally, type 3 corresponds with the remaining individuals with very high reports (this time on average federal tax reports, since the sample selection cut removed high marginal rate reporters), as $\beta_{0,A}$ is 80.5. The other coefficients exhibit the exact opposite pattern as expected from the "rational" or "correct" type

hypothesized: the individual's average rate effects their marginal report while their marginal rate significantly, at the 10% level, affects their average report, but the "correct" rate to use is insignificant in each case. There is a 1% probability of being classified as this type in the data. Looking at these results in total, it appears that the data lends itself to being classified into types based on the general level of tax rates reported, instead of based on the theoretical models presented earlier.

2.5 Subsample analysis

While there are a wide variety of actual federal rates observed in the sample (over 27 rates), most of these apply to a very small number of individuals. In particular, many rates apply for small (measured by income range) notches in the federal tax codes created by features such as clawbacks in deductions or phase-out of credits. Other rates cover a limited number of respondents with either very high or very low tax rates. This feature of the data creates two potential problems for analysis. The first problem is measurement error in the income variable which may place the household incorrectly in a small notch that is a tax rate that the household does not truly face. The second problem is the limited number of observations in regions (such as high or low income) who may perceive taxes in a very different way than the broader sample, but for whom we have limited data. Therefore, we focus an additional analysis on the sample in the two federal income-tax brackets for which we have the most data, those who are in the 15% bracket and the 25% bracket, containing 1,078 and 467 respondents, respectively. These tax brackets are adjacent in the federal tax schedule (individuals move from the 15% bracket to the 25% bracket). Focusing on these individuals allows us to have the power to examine the effects of underlying rate changes in a way that most of the other

rate changes we observe do not.

2.5.1 Reduced-form analysis

We begin by examining how individuals report their marginal tax rates across the two income-tax levels. First, we examine how errors in reports correspond with the individual being in each of four regions of the tax schedule: in the middle of the 15% federal tax rate, in the middle of the 25% federal tax rate, near the boundaries of the 15% tax rate, or near the boundaries of the 25% tax rate. Here, as before, we define the boundaries as being in a range where a change of 10% of household income would put the individual in a new tax rate. The regression specification is:

$$(\tau_F^M - \tau_F^{*M}) = \beta_0 + \beta_1 I_B + \beta_2 I_{F25} + \beta_3 I_B I_{F25}$$

τ_F^{*M} is the actual federal marginal tax rate, I_B is an indicator for the respondent being within one of the tax schedule boundaries, and I_{F25} is an indicator for being in the 25% federal tax bracket.

Looking at the results in column 1 of Table 2.8, we find only two statistically significant results. The constant of 2.34 indicates that respondents in the middle of the 15% bracket are, on average, overestimating their federal income tax by roughly 2.3 percentage points. The coefficient estimate for being in the 25% bracket of -6.62 indicates respondents in this tax range underestimate their marginal tax rate by roughly 4.3 percentage points. Interestingly, this means that, on average, the respondents across the two federal-tax rates, which differ by 10 percentage points, only change the average rate reported by less than 2 percentage points:

respondents in the 15% bracket report an average of a 17% tax rate, while those in the 25% bracket report an 21.7% tax rate. The effect of being near a boundary is statistically insignificant.

However, there might be some difference between being near the upper boundary of a tax rate relative to the lower boundary. Therefore we next we split the income region into six areas. We accomplish this by splitting each boundary region into the lower (near the step to the lower marginal tax rate) and the upper (near the step to the next higher marginal tax rate). This allows us to detect if there are differences in perceptions arising from being just below a kink, compared to being just above a kink. The estimates from this regression are presented in column 2 of Table 2.8. In this regression, we once again find that people in the middle of the 15% bracket tend to overestimate their taxes and people in the middle of the 25% bracket tend to underestimate their taxes. However, this specification allows us to detect statistically significant differences in reporting behavior near the kinks. Examining the indicator variable for each region, we find a distinct pattern in reporting as we move from left to right across the tax schedule (in the direction of increasing incomes and increasing tax rates). Individuals closest to the lower bound of the 15% bracket (i.e. closest to the 10% federal tax bracket), underreport their tax rate by 1.7 percentage points. As per above, those in the middle of the 15% bracket overreport their tax rate by 2.3 percentage points. As income increases, and respondents near the upper bound of the 15% bracket, they overreport by 5.7 percentage points. While at the lower bound of the 25% bracket respondents underreport by 5.3 percentage points. In the middle of the bracket they underreport by 4.3 percentage points. Finally, at the upper edge they under report by 2.2 percentage points. Combining these, we find a generally steady increase in the average rate reported from 13.3 to 17.3 to 20.7 to 19.7 to 20.7 to

22.8. Interestingly, the only place the respondents' rate has actually increased is in between the average report of 20.7 and 19.7, where it has jumped by 10 percentage points. This pattern seems somewhat consistent with a notion of average-tax-rate reporting, combined with some initial overreporting of estimates.

Table 2.8: Errors in federal report by location in income tax bracket

I_{F25}	-6.62***	
	(0.656)	
I_B	0.028	
	(0.670)	
$I_{F25} * I_B$	-0.588	
	(1.148)	
I_{15}		4.07***
		(0.914)
I_{UB15}		7.41***
		(1.14)
I_{LB25}		-3.60***
		(1.22)
I_{25}		-2.61***
		(1.00)
I_{UB25}		-0.53
		(0.794)
Constant	2.342***	-1.73**
		(0.85)
R^2	0.10	0.12
N	1,545	1,545

Notes: Regresses error in federal marginal rate report by location in the federal income tax schedule. I_{F25} is an indicator for being in the 25% federal tax bracket, I_B is an indicator for being with 10% of current income of a kink in the tax schedule. and I_{LB25} and I_{UB25} are indicators for being within 10% of current income from the lower boundary and upper boundary of the 25% tax rate, while I_{25} is an indicator for being in the middle of the 25% rate (i.e. not within 10% of the boundary). Similarly for I_{15} and I_{UB15} in the 15% rate.

2.5.2 Model of perception

We next estimate the mixture model of perception in this sample. In order to accommodate the fact that there are now only two actual federal marginal tax

rates, we no longer include marginal tax rate as a continuous variable and instead include an indicator for being in the 25% federal income tax bracket. The results of this model are presented in Table 2.9.

Table 2.9: Mixture model results, subsample

Type 1:					
Dep var: $\hat{\tau_F^M}$			Dep var: $\hat{\tau_F^A}$		
	I_{25F}^M	11.15***		I_{25F}^M	7.11**
		(1.33)			(1.18)
	τ_F^A	0.47***		τ_F^A	0.36***
		(0.13)			(0.14)
	Constant	8.35***		Constant	6.72***
		(1.03)			(1.07)
Type 2:					
Dep var: $\hat{\tau_F^M}$			Dep var: $\hat{\tau_F^A}$		
	I_{25F}^M	-16.28***		I_{25F}^M	-8.39***
		(1.47)			(1.45)
	τ_F^A	0.28**		τ_F^A	0.25
		(0.14)			(0.16)
	Constant	22.15***		Constant	14.7***
		(1.20)			(1.24)
Type 3:					
Dep var: $\hat{\tau_F^M}$			Dep var: $\hat{\tau_F^A}$		
	I_{25F}^M	-26.14***		I_{25F}^M	11.18
		(6.46)			(7.71)
	τ_F^A	1.85***		τ_F^A	-0.31
		(0.68)			(0.80)
	Constant	5.58		Constant	85.98***
		(5.30)			(6.22)
		$N = 1,545$			

Notes: Recreates Table 2.6 with the two rate subsample.

The estimates for type 1 are broadly consistent with the general findings from the full-sample model. Coefficients for both average and the current rate indicator are positive and statistically significant, indicating that both are influencing both types of individual reports in this type. In this type we see the effect of changing tax brackets in the individual reporting of marginal rates is 11 percentage points

(compared with the true effect of a 10 point increase in the underlying federal rate). These individuals appear to be approximately calibrated to the increase in marginal federal reports. Similarly, the effect of changing brackets on average reports is 7 percentage points, which is in line with the fact that average rates increase slower than marginal rates. The estimates for type 2 are more puzzling, as now average rates have similar effects on reporting (although not statistically significantly affecting the average reports), but now the indicator has a large and significant negative effect on reporting for both marginal and average rates (-16.3 and -8.4 percentage points, respectively). This change is coupled with an increase of the constant terms in both regressions, up to 22.2 and 14.7 percentage points, respectively. Consequently, this type contains those who are overreporting if they are in the 15% tax bracket and underreporting if they are in the 25% tax bracket. Using the median tax rate in each group, the average marginal reports would be 24.1% and 9.8% for those in the 15% and 25% marginal brackets. As in the previous estimation, the third type contains those with very high average reports. As reported in Table 2.10, the majority of people are in type 1, with about 58% probability of assignment, type 2 has a 41% probability and type 3 remains at just over 1%.

Table 2.10: Latent class probabilities, subsample

Class	
1	0.579
	(0.031)
2	0.408
	(0.031)
3	0.013
	(0.003)

Notes: Posterior probabilities of class membership from mixture model estimated in Table 2.9.

2.6 Conclusion

This paper has provided experimental evidence on income-tax perceptions. The findings are mixed, much like the existing literature. We find some evidence of individuals understanding progressivity (although with much too much curvature in the reported schedule relative to the true schedule) and some evidence that reports behave more like averages than marginals (in the patterns seen in reporting errors). This paper provides the first evidence that demonstrates individuals seem to have similar perceptions of state and federal income tax schedules, having broadly consistent patterns across all metrics examined, particularly in states with progressive income tax structures.

This paper was unable to estimate a direct model of state perceptions and it appears that much more data for a single state (or similar states) would be required to do so. This would be an interesting next step for the literature, although the initial evidence leads us to believe the findings would likely be similar to those in federal tax reports.

Another question raised by this paper is what effect does individuals' "over"-perceiving progressivity have on labor-supply choices. It is possible that individuals are less incentivized to earn higher wages or work more hours due to the mistaken sense that taxes will go up quickly and reduce their net increase in earnings. This finding could be further exacerbated by the fact that individuals also perceive the tax rate to decrease more quickly than it does below them, meaning the perceived decrease in post-tax income from working less would be smaller than the actual decrease. The combined effect of these perceptions is that individuals perceive an overly smooth marginal tax schedule with more frequent changes than actually exist. This mismatch between perception and reality could explain the empirical

finding of a lack of bunching around actual kinks in the tax schedule (Saez, 2010). The mismatch in perceptions and reality would also bias estimates of the elasticity of taxable income using such bunching estimators as there is a disconnect between the rate changes individuals perceive and those that actually exist.

Additionally, while taxes are known to distort (and reduce) labor-supply decisions, misperceptions of the tax schedule are, as of yet, generally unstudied in their labor-supply effect. The effect of "over-progressivity" on potentially reducing labor supply is particularly notable when combined with this paper's findings when looking at the 15% and 25% federal-rate subsamples. There we found that individuals in the lower tax rate not only believed their tax rate was going up faster than it did, they also believed their rate was already higher than it actually was (as compared to those in the higher tax rate who actually underestimated their tax rate). This creates two forces that would tend to bias this group's labor supply downward. This paper does not have a sufficient sample in the lowest tax rate (10%) to estimate it, but if these twin trends continue then the group of extremely low earners may have particularly large (and incorrect) incentives not to work due to errors in perception. This may be an area of particular interest for future research and policymakers due to the increasing trend of tying state and federal aid to labor market participation.

CHAPTER 3

DIFFICULTY TO REACH RESPONDENTS AND NONRESPONSE BIAS: EVIDENCE FROM LARGE GOVERNMENT SURVEYS (WITH ORI HEFFETZ)

3.1 Introduction

To what extent do survey-based estimates depend on the difficulty of reaching respondents? The answer can hint at how cautious one should be when making population-wide inferences from surveys with low response rates. If within a survey sample, after controlling for other observables, outcomes are systematically different across easy-to-reach and hard-to-reach respondents, then one may question the routinely made assumption that nonrespondents—those out of sample who are, effectively, the hardest to reach—are similar to the average within-sample respondent. In other words, within-sample comparisons across difficulty-of-reaching groups can help assess the assumption of (conditional) random selection into the survey sample. This paper reports findings from such within-sample comparisons. Its goal is to help in assessing how sensitive population-wide estimates of important outcomes are to survey response rates and to assumptions regarding nonrespondents.

Our (purely empirical) investigation proceeded in three steps. First, we identified three large and widely used government surveys that report the number of phone or in-person visit attempts made in the course of reaching respondents: the Census/BLS's Current Population Survey (CPS), CDC's Behavioral Risk Factor Surveillance System (BRFSS), and BLS's Consumer Expenditure Survey (CEX). The three show a wide range of response rates, which have all been generally de-

clining over the years, averaging in our data from just under 45% in the BRFSS to around 70% in the CEX and to just over 90% in the CPS. We note that these are the only three datasets that we investigated, and we chose in advance to report our findings from all three.

Second, within each dataset, we sought to identify one or two key outcomes to analyze. We were guided by our goal to focus on the outcomes of most interest to researchers, policymakers, and the public. Our search resulted in choosing four key outcomes: the labor force participation rate and the unemployment rate from the CPS, obesity prevalence from the BRFSS, and total household expenditures from the CEX. All four are national statistics that are closely watched, analyzed, and discussed in the academic literature, business world (or, in the case of obesity, health-policy world), and popular media. Importantly, unlike opinion-poll and social-survey outcomes such as voter intentions or consumer and investor confidence—outcomes that some researchers fundamentally mistrust—our four key outcomes are regarded by many as (more or less) objective indicators, based on survey questions designed to elicit factual reports rather than perceptions, feelings, or convictions.[1]

Third, based on the number of contact attempts made to each respondent, we divided each dataset into three or four difficulty-of-reaching groups, as similar in size as we could. We then compared, within each dataset, easy-to-reach respondent groups versus hard-to-reach groups with regard to outcome averages (both unadjusted and adjusted for observables) and to cross-demographic-group differences.

[1]While not a key outcome of our paper, we also analyzed a life satisfaction question asked in the BRFSS, and used it to probe the robustness of past research on subjective well-being data (see Section 3.3.1).

We find strong and robust differences between easy-to-reach and hard-to-reach respondents in all four primary outcome variables. Briefly, the labor force participation rate (65.1% overall in our 2012–2013 CPS sample) is 4.9 percentage points lower among the easiest-to-reach respondents (reached in a single visit attempt) than among the hardest-to-reach respondents (reached in 3 or more attempts), after controlling for demographic differences across difficulty-of-reaching groups. Similarly, the unemployment rate (7.6% in our sample) monotonically decreases 1.5 percentage points from easiest- to hardest-to-reach; the obesity rate (28.4% in our 2012 BRFSS sample) monotonically decreases 3.1 points from easiest (1 call attempt) to hardest (7+ attempts); and average log quarterly household expenditures ($9,459 in our 2008–2013 CEX sample, exponentiated back to dollars) increase by $465 from easiest- (1 contact attempt) to hardest-to-reach (5+ attempts). In addition, for labor force participation and for obesity—but not for the other two outcomes—the male-female gap and other cross-demographic-group gaps consistently shrink or increase with difficulty of reaching.

Overall, our analysis reveals a consistent picture: in our data, difficulty-of-reaching is strongly correlated with important outcomes of interest, even after controlling for the main observables that typical weighting schemes are based on. How important is this finding?

In principle, the finding of systematic differences between easy- and hard-to-reach respondents is not, by itself, necessarily worrisome. As long as survey nonrespondents are randomly selected (unconditionally, or on observables) from the population-representative sample targeted by the survey, sample averages (unconditional, or conditional on observables) could be made generalizable to the population. But a difficult-to-reach respondent in one survey could be a nonrespondent

in another survey that had a higher nonresponse rate due to time, budget, or other constraints. By the same token, nonrespondents in a given survey can be viewed simply as respondents so difficult that they remained out of reach. Indeed, the relatively high response rates in surveys such as the CPS suggest that additional effort and resources can bring many difficult (non)respondents into the sample. From this perspective, a finding of within-sample differences across difficulty-of-reaching groups challenges the random-selection assumption: if difficulty of reaching is correlated with outcomes, the likelihood of nonresponding may also be correlated with outcomes. Moreover, if in-sample trends in differences across difficulty-of-reaching groups extend to (out-of-sample) nonrespondents then not only are nonrespondents different from the average respondent; their average outcomes are not even within the range of average outcomes observed within sample.[2]

If we could somehow observe nonrespondents' outcomes—e.g., by matching respondents and nonrespondents with administrative measures of the survey outcomes we investigate—we could directly examine whether the difficulty-outcome trends we find extend to nonrespondents, and hence whether these trends are indeed evidence of nonresponse bias. (We could then also directly investigate another important question that our paper cannot examine: that of survey measurement error.) In practice, such direct tests are impractical for the same reason that the government surveys we study are so widely used in the first place: nationally representative administrative datasets containing "true" measures of the outcomes we study are not readily available.[3]

[2]Heffetz and Rabin (2013, p. 3007) provide a step-by-step numerical example that illustrates this point.

[3]In fact, it is not even clear that administrative data could exist for outcomes such as labor force participation and unemployment rates as defined and measured in the CPS, since they require reports of job search activity. On the other hand, in certain other special cases researchers *are* able to match administrative and survey data. We are aware of two papers—Lin and Schaeffer (1995) regarding child-support awards and payments, and Kreuter et al. (2010) regarding welfare receipts and other outcomes—that investigate difficulty trends where true outcomes are known

While we cannot use our data to directly assess whether the within-sample trends we find extend to nonresponders, the wide range of response rates in the three datasets we study provides indirect, circumstantial evidence. In the CPS, with a nonresponse rate just under 10%, we do observe many who would be nonrespondents in another survey. The consistent in-sample trends we find in the CPS may hence tentatively provide some rationale for out-of-sample extrapolation of in-sample trends in higher-nonresponse-rate surveys—while acknowledging, of course, that such cautious rationale may apply to certain types of nonresponders more than to others. More generally, our finding of consistent trends in three surveys with, respectively, around 55%, 30%, and 10% nonresponse rate suggests that nonrespondents may on average look more like very difficult-to-reach respondents than like the average in-sample respondent. (Indeed, as hinted above, the very coexistence of such a wide range of nonresponse rates may in itself suggest it.)

Beyond these cautious arguments, the assumption that nonrespondents look like the in-sample average is simply hard to defend when said average is a moving target, changing systematically as increasingly difficult respondents are added to the sample. Moreover, when tracking outcomes over time, that assumption may *differentially* impact estimates of both long-run trends—as response rates slowly decline—and shorter-run fluctuations—as the outcome-difficulty gradient could itself vary with the outcome (we discuss such evidence in our concluding section). In the typical case—such as ours—where researchers do not know much about nonresponders, one may therefore view a finding of large within-sample difficulty-of-reaching differences as a reason for concern, the more so the higher is the nonresponse rate.

for respondents and nonrespondents. As we discuss in detail in Section 3.5, the former finds that the difficulty trends among respondents mostly do not extend to nonrespondents, but its survey context and difficulty measure are rather different from ours, while the latter, whose setup and difficulty measure are arguably more comparable to ours, finds that they do.

The rest of our paper proceeds as follows. In Section 3.2 we study the CPS. We find large cross-difficulty-of-reaching differences in the raw (i.e., unadjusted) subsample means of our two key outcomes. The labor force participation rate climbs from 63.0% [SE 0.1] in the easiest-to-reach group to 72.3% [0.2] in the hardest-to-reach group—an increase of over 9 percentage points—while the unemployment rate declines from 8.1% [0.1] to 6.7% [0.2]. Of course, if these differences in outcomes were explained entirely by demographic differences between the easy- and hard-to-reach respondents then—assuming the demographic composition of the population is known from outside sources—applying the correct weights would yield unbiased population estimates. To show that this is not the case, at least not in a simple way, we compare adjusted means (controlling for observables including age, sex, race, education, and others) and still find that the adjusted labor force participation rate increases from 64.1% [0.1] to 69.0% [0.3], and the adjusted unemployment rate decreases from 8.0% [0.1] to 6.5% [0.2].

We also find that differences in labor force participation across age, sex, and other demographic groups change systematically with difficulty of reaching, albeit less dramatically. Finally, in robustness analysis we find that our CPS findings are even stronger when limiting the sample to self reports, for whom our difficulty-of-reaching measure is likely cleaner than for proxy reports; and are weaker when limiting the sample to telephone-completed interviews, for whom the measure is likely noisier.

In Section 3.3 we study obesity prevalence in the BRFSS. Here the uncontrolled means decrease from 29.4% [SE 0.1] to 27.1% [0.1] from the easiest- to the hardest-to-reach groups, and the adjusted means decrease from 29.7% [0.1] to 26.6% [0.2]. We also find some evidence of cross-demographic-group differences in obesity that

change with difficulty of reaching. For example, among BRFSS's easiest-to-reach respondents, males' adjusted obesity rate is indistinguishable from females' rate (male−female diff: 0.4% [SE 0.3]), however in the hardest-to-reach group males are a further 1.8 [0.4] percentage points more likely to be obese than females (diff: 2.2%). We illustrate the potential practical implications of these findings using a simple extrapolation. We also explore the relationship between difficulty of reaching and the self-reported height and weight variables that underlie the BRFSS obesity measure. Finally, we show that using the survey weights that the BRFSS recommends for making population-wide inferences does not change our conclusions.

In Section 3.4 we analyze log quarterly expenditures of CEX households, again finding a notable trend across difficulty-to-reach groups. Transformed back into dollars, unadjusted means rise from $8,225 [SE $74] to $9,990 [64], and adjusted means rise from $9,120 [59] to $9,585 [43]—a 5% increase—from easiest to hardest. We again demonstrate the robustness of our estimates in several ways. We also analyze separately total food and total health expenditures and find that they increase and decrease, respectively, with difficulty of reaching.

In Section 3.5 we draw connections with previous work that investigates the relationship between difficulty of reaching and outcomes. To our knowledge, such work is almost non-existent in economics. We review work—mostly in statistics and survey methodology—that has been exploring the potential use of difficulty-of-reaching measures and other paradata (data about how the data were collected) for making inferences regarding nonrespondents. Importantly, we review evidence that suggests that higher response rates do not always reduce overall bias: while potentially reducing unit nonresponse bias, they may increase item nonresponse

bias or measurement error if, for example, the harder to reach are less likely to accurately answer certain survey questions. We also review evidence on the quality of number-of-contact-attempts paradata as difficulty-of-reaching measures.

Our paper's closest predecessor is Heffetz and Rabin (2013), whose closest predecessor is Curtin et al. (2000). These earlier papers study outcomes from the University of Michigan's Surveys of Consumers: self-reported happiness, and the Index of Consumer Sentiment (ICS), respectively. Both papers find that conclusions regarding outcome variables depend on the difficulty of reaching respondents. Indeed, the outcome measures they study—designed to elicit mostly unverifiable emotions, attitudes, and beliefs—may be directly affected by the momentary context in which they are asked and hence may be more prone to difficulty-of-reaching differences. (For example, busy people may only be reached late in the day, when they are more pessimistic; or over the weekend, when they are more optmistic.) In contrast, our main outcomes of interest—labor force participation, unemployment, obesity, and expenditures—are measures that are designed to reflect verifiable states that do not change moment to moment, and as such are supposedly unaffected by the momentary situation related to the availability (or busyness) of respondents. A main contribution of our paper is to document that conclusions regarding such outcomes *also* depend on the difficulty of reaching respondents, in a way that simple reweighting schemes cannot correct.

We conclude in Section 3.6, where we discuss the practical implications of our findings—both for users of the specific outcomes we study and more broadly for designers and users of other surveys. As a concrete application, we plot unemployment data over two decades (1994–2013) and demonstrate that cyclical unemployment fluctuations have been more extreme among easy than among difficult

respondents.

3.2 CPS: Labor Force Participation and Unemployment

3.2.1 Data

The Current Population Survey (CPS) is a monthly survey conducted by the Bureau of Labor Statistics (BLS) and the U.S. Census Bureau. Among its many uses, it provides closely watched labor force statistics for the U.S., such as the labor force participation and unemployment rates. The survey consists of a rotating panel of households. A participating household provides data for four consecutive months, is not contacted for eight months, and then provides data for four more consecutive months. The eight interviews are referred to as month-in-sample (MIS) 1 through 8. The CPS samples addresses, so each household's first interview (MIS 1) must begin with a personal visit by a field interviewer, but it can be completed by phone if requested by the respondent. The majority (around 85%) of the remainder of the interviews are conducted via telephone, with the exception of MIS 5, the one following the eight-month break, which is also usually completed in person.

Within each household, the interviewer attempts to interview the "most knowledgeable" member, often the owner or renter of the unit (CPS, 2015). During each interview the respondent is asked to help create a roster of eligible household members and answer questions to determine, for each member, whether she is: employed, unemployed, or out of the labor force. Specifically, for each civilian member of the household who is 15 or older, a series of questions is asked to determine whether or not she was employed in the interview month's reference week

(almost always the week containing the 12th of the month). If she was, the CPS codes her as employed. Otherwise, the interview determines, among other things, whether or not she was actively looking for work in the past four weeks and was available for work during the reference week. If she was, the CPS codes her as unemployed. Otherwise—i.e., she neither had a job (employed) nor was actively looking for one (unemployed)—she is coded as out of the labor force.

We analyze CPS data from January 2012 to December 2013. The CPS response rate in each of these 24 months is between 89% and 91%, and is 90.1% overall for the period.[4] Our analysis sample consists of all 307,603 MIS 1 observations that the BLS would include when calculating the labor force participation and unemployment rates—i.e., civilians aged 16 and up with complete labor force records.[5] The labor force participation rate is defined as the share of these who are in the labor force (i.e., they are either employed or unemployed); the unemployment rate is defined as the share of those in the labor force who are unemployed.

Finally, our difficulty-of-reaching measure is a variable recording the "number of actual and attempted personal contacts." It is a noisy measure. Importantly, it counts only personal visit attempts, so it likely understates the difficulty of reach-

[4]These are calculated from the CPS monthly files: the number of households with a partial or complete interview is divided by the number of eligible households in the month. This calculation corresponds to the American Association of Public Opinion Research (AAPOR) Response Rate #2, #4, and #6 (AAPOR, 2016). (The three response rate definitions differ in how households of unknown eligibility are treated, but the CPS contains no such households because interviewers classify the eligibility of each address.) Krueger et al. (2017, Figure 4) show that nonresponse rate in the CPS has been generally increasing, from slightly above 4% in 1990 to roughly 8% in 2010, and then steeply climbing to nearly 11% by 2014.

[5] We restrict our analysis to MIS 1 data for two reasons. First, the number-of-contact-attempts variable only counts *in-person* contact attempts, so it is mostly non-zero only for MIS 1 (where 91.9% of the observations in our analysis sample are non-zero) and, to a lesser extent, for MIS 5 (76.1% non-zero). Second, the CPS is subject to rotation group bias, i.e., some of its outcomes vary systematically by MIS (for recent evidence see and Dixon (2013) and Krueger et al. (2017)); focusing on a single MIS avoids this confounding bias. (While the sources of this bias are still being investigated, Krueger et al. (2017) find suggestive evidence that the unemployment rate calculated from early interviews—MIS 1, 2, 3, and 5—is a stronger predictor of other measures of economic slack than that calculated from later interviews.)

ing respondents who were also contacted by the telephone.[6] In addition, since it is reported by the interviewer (and not by an automatic system, as in some telephone surveys), it may be affected by intentional or unintentional misreporting.[7] Indeed, 24,807 observations (8.1% of our analysis sample) have nil contact attempts reported, and we do not know how difficult to reach they were; in the tables below we classify their number of contact attempts as "None Reported," or "NR." The remainder of the observations have 1–9 (top coded) contact attempts recorded. We classify them into three categories: 1 attempt (64.3%), 2 attempts (17.0%), and 3 or more attempts (10.7%).

3.2.2 Analysis by Difficulty of Reaching

Sample composition

Table 3.1 reports basic demographics for our sample. Each of its first four columns is based on a single number-of-contact-attempts category (1, 2, 3+, and NR); the fifth column is based on the entire sample (All). The first three columns show that on some demographics the sample's composition changes systematically with number of attempts. Notably, the young are harder to reach than the old: those aged 20–39 comprise 30.7% of the 1-attempt sample compared with 37.0% of the 3+-attempts sample, while those aged 65 and up comprise 20.2% and 12.5% respectively. (In both cases, the 2-attempt percentages lie in between.) On the other hand, women and men are of overall similar difficulty of reaching: the share

[6] 52,624 of the observations in our sample (17.1%) are known to have been conducted at least in part via telephone, and we also analyze them separately below. Another 6,398 observations (2.1%) have a missing value for interview mode.

[7] While not explicitly stated in the data dictionary, it was confirmed to us by researchers at the BLS that this variable, as well as the mode variable, are manually entered by the interviewer.

of women, 52.2% overall, does not vary much with number-of-attempts category. We also note that demographic-group shares in the NR column are in some cases inside, and in other cases outside (below or above), the range of shares in the three leftmost columns. This makes it difficult to hypothesize about the difficulty-of-reaching of NR respondents, and we have little to say about them in the rest of this section; we include them in the tables only for completeness.

Looking at our variables of interest at the bottom of the table, the labor-force-status composition of the sample changes dramatically with contact attempts: a sharp increase in the share employed, from 57.9% (1 attempt) to 67.5% (3+ attempts), is mirrored almost entirely by a sharp decrease in the share not in the labor force, from 37.0% to 27.7%, while the share unemployed is pretty stable and hovers around 5%. In other words, as contact attempts increase, household members are more likely to be recorded as employed and less likely to be recorded as not employed who are not looking for employment. As a result, and as discussed in the introduction, our two key outcomes—the labor force participation rate and the unemployment rate—increase, and decrease, respectively, with contact attempts. In the rest of this subsection we show that this increase and decrease are not eliminated by controlling for the (first order) changes in demographic composition noted above.

Table 3.1: CPS Demographics

Attempts		1	2	3+	NR	All
Age:	16–19 (%)	6.4	7.0	7.1	6.1	6.5
		(0.1)	(0.1)	(0.1)	(0.2)	(0.0)
	20–39	30.7	33.4	37.0	28.4	31.6
		(0.1)	(0.2)	(0.3)	(0.3)	(0.1)
	40–49	16.5	18.1	18.8	17.3	17.1
		(0.1)	(0.2)	(0.2)	(0.2)	(0.1)
	50–64	26.2	25.7	24.7	28.9	26.2
		(0.1)	(0.2)	(0.2)	(0.3)	(0.1)
	65 and up	20.2	15.8	12.5	19.3	18.6
		(0.1)	(0.2)	(0.2)	(0.3)	(0.1)
Children in household		25.5	28.4	28.5	23.4	26.2
		(0.1)	(0.2)	(0.2)	(0.3)	(0.1)
Female		52.2	52.0	52.0	52.8	52.2
		(0.1)	(0.2)	(0.3)	(0.3)	(0.1)
Educ:	Less than high school	14.9	14.7	14.1	11.1	14.5
		(0.1)	(0.2)	(0.2)	(0.2)	(0.1)
	High school	30.2	29.1	28.0	27.1	29.6
		(0.1)	(0.2)	(0.2)	(0.3)	(0.1)
	Some college or tech. school	27.3	28.0	28.3	27.0	27.5
		(0.1)	(0.2)	(0.2)	(0.3)	(0.1)
	College graduate	27.6	28.2	29.6	34.8	28.5
		(0.1)	(0.2)	(0.3)	(0.3)	(0.1)
Race:	White	82.9	81.1	78.8	84.8	82.3
		(0.1)	(0.2)	(0.2)	(0.2)	(0.1)
	Black	9.7	10.1	11.5	8.6	9.8
		(0.1)	(0.1)	(0.2)	(0.2)	(0.1)
	Asian	4.4	5.6	6.4	4.4	4.8
		(0.0)	(0.1)	(0.1)	(0.1)	(0.0)
	Other	3.1	3.2	3.3	2.3	3.1
		(0.0)	(0.1)	(0.1)	(0.1)	(0.0)
L.F.P.:	Employed	57.9	62.5	67.5	63.5	60.2
		(0.1)	(0.2)	(0.3)	(0.3)	(0.1)
	Unemployed	5.1	5.2	4.8	4.0	5.0
		(0.0)	(0.1)	(0.1)	(0.1)	(0.0)
	Not in the labor force	37.0	32.3	27.7	32.5	34.9
		(0.1)	(0.2)	(0.2)	(0.3)	(0.1)
Labor force participation		63.0	67.7	72.3	67.5	65.1
		(0.1)	(0.2)	(0.2)	(0.3)	(0.1)
Unemployment rate		8.1	7.6	6.7	5.9	7.6
		(0.1)	(0.1)	(0.2)	(0.2)	(0.1)
Median number of attempts (known)		1	2	3	n/a	1
Observations		197,751	52,275	32,770	24,807	307,603

Notes: Source: Current Population Survey, Jan. 2012–Dec. 2013. Sample: all MIS 1 observations who are qualified to be in the civilian labor force and gave enough employment information to be classified. All figures (and standard errors) reflect proportions within each column's difficulty-of-reaching category, except for those for unemployment rate (which are calculated as described in text) and number of attempts (which report medians). NR: No reported contact attempts.

Labor force participation

Table 3.2 reports our main labor force participation results. Since the table's structure is shared by all other main-results tables in the rest of this paper (with unemployment, obesity, and expenditures as dependent variables), as well as with many appendix tables, we describe it below in some detail. We also note here that our main findings are not affected more than trivially by including a larger set of age indicators and interactions or, for our *binary* dependent variables (labor force participation, unemployment, and obesity), by replacing the OLS specification in our main-results tables with probit or logit.[8]

Table 3.2's four columns report results from a *single* OLS regression. The dependent variable is a 0/1 labor force participation indicator. The regressors are sets of demographic indicators (those reported in Table 3.1, plus unreported indicators for marital status (6 categories), household size (5), state (51), urban/rural (3), interview month (12), interview year (2), and a constant); a set of difficulty-to-reach-category indicators; and a full set of interactions of the difficulty indicators $\times$ all demographic indicators (including those unreported). Panel A reports the estimated coefficients: the first column reports coefficients on the demographic indicators for the base (omitted) difficulty-to-reach category (1 contact attempt), and the other columns report the coefficients on the demographic indicators interacted with each of the three other difficulty categories (2, 3+, and NR). Notice the reported 4×13 coefficients are mechanically identical to those one would get from estimating a separate regression of the dependent variable on the set of demo-

[8]Specifically, all the OLS-adjusted means discussed below remain within their reported standard errors (and are often within our tables' rounding error) when including a set of ten rather than five age categories, as well as all age-gender, age-education, and gender-education interactions; and all the probit/logit-adjusted means are within 0.13 percentage points of the OLS-adjusted means discussed below (indeed, most are exact matches, given rounding error), with the interaction coefficients showing the same patterns and statistical significance.

Table 3.2: Labor force participation

Attempts	1	2	3+	NR
	A: Regression with interactions			
	Base	Interactions		
Age: 16–19	-0.263***	-0.025**	-0.035**	-0.047***
	(0.006)	(0.012)	(0.015)	(0.018)
20–39	-0.011***	-0.009	-0.003	-0.024***
	(0.003)	(0.006)	(0.006)	(0.008)
50–64	-0.113***	0.018***	0.049***	0.028***
	(0.003)	(0.006)	(0.007)	(0.008)
65 and up	-0.574***	0.016*	0.069***	0.032***
	(0.004)	(0.008)	(0.011)	(0.011)
Children in household	0.039***	0.017***	0.035***	0.026***
	(0.003)	(0.007)	(0.007)	(0.009)
Female	-0.103***	-0.002	0.010**	0.007
	(0.002)	(0.004)	(0.005)	(0.005)
Educ: Less than high school	-0.136***	0.008	0.001	-0.020*
	(0.003)	(0.007)	(0.009)	(0.012)
Some college or tech. school	0.029***	-0.004	-0.010	-0.004
	(0.003)	(0.006)	(0.007)	(0.008)
College graduate	0.098***	-0.014***	-0.025***	-0.001
	(0.003)	(0.005)	(0.006)	(0.007)
Race: Black	-0.026***	-0.004	0.001	-0.016
	(0.004)	(0.008)	(0.009)	(0.011)
Asian	-0.055***	-0.000	0.013	0.011
	(0.005)	(0.010)	(0.012)	(0.015)
Other	-0.037***	-0.006	0.015	0.047**
	(0.006)	(0.013)	(0.016)	(0.020)
Constant	0.782***	0.016	0.105***	0.107***
	(0.011)	(0.024)	(0.027)	(0.038)
	B: Adjusted means			
Labor force participation	0.641***	0.662***	0.690***	0.665***
	(0.001)	(0.002)	(0.003)	(0.003)

Notes: Source: Current Population Survey, Jan. 2012–Dec. 2013. $N = 307,603$ (1 attempt: 197,751; 2: 52,275; 3+: 32,770; None Reported: 24,807). $R^2 = 0.29$. The table reports estimates from a single OLS regression. Dependent variable: 0/1 labor force participation indicator. See page 85 for a full explanation of table structure. Panel A: estimated coefficients from a fully interacted regression: each regressor is interacted with each difficulty-to-reach category (omitted category: 1 attempt). Regression also includes non-reported indicators (and their interactions) for marital status (6 categories), household size (5), state (51), urban/rural (3), interview month (12), and interview year (2); see appendix note B.1. Standard errors, clustered at the household level, in parentheses. Panel B: adjusted means, calculated from panel A regression. *** p<0.01, ** p<0.05, * p<0.1.

graphics (and no interactions) within each of the four difficulty categories (that is, four separate regressions), and then subtracting the coefficients in the 1-attempt regression from those in each of the other three regressions.

Panel B reports adjusted means for the four difficulty-to-reach categories (calculated from the regression coefficients estimated in Panel A). Intuitively, the adjusted means are calculated as follows.[9] For each observation, one calculates the dependent variable's predicted value four times, using that observation's true values for all the independent variables except for the set of difficulty-to-reach indicators, which are changed to indicate 1 attempt for the 1-attempt adjusted mean, are changed to indicate 2 attempts for the 2-attempts adjusted mean, and so on. The adjusted means row then reports these predicted values averaged across all the sample's observations. The four adjusted means are hence the average predicted value of the dependent variable in four hypothetical samples. Each of the four samples is identical to our actual (full) sample except that the number-of-contact-attempts category is hypothetically set to 1, 2, 3+, or NR, respectively, for all of that sample's observations.

We start with the bottom line: the adjusted means in Panel B are 64.1% (1), 66.2% (2), and 69.0% (3+)—a total increase of 4.9 percentage points, with each of the three point estimates statistically different from the others. In other words, the differences in composition across the difficulty-to-reach categories explain less than half of the 9.3-point raw increase from Table 3.1.

Looking at panel A, the fully interacted regression shows that not only are respondents in different difficulty categories predicted to have different labor force participation rates even when they are otherwise identical (on observables) to the

[9] In practice, we use the STATA 14.1 command "margins," which also computes standard errors using the delta method. Our intuitive exposition here draws in part on Williams (2011).

entire sample, but furthermore these predicted differences interact with demographics. For example, teenagers (aged 16–19) are 26.3 percentage points less likely to participate than those aged 40–49 among 1-attempt respondents but are $26.3 + 3.5 = 29.8$ points less likely to participate among 3+-attempts respondents. In contrast, those 65 and up are 57.4 points less likely to participate with 1 attempt, but are 50.5 points less likely to participate with 3+ attempts, a (highly significant) 6.9 point reduction in the difference. As to other age-group coefficients, that on ages 20–39 seems relatively stable, while that on ages 50–64 shrinks from 11.3% (1 attempt) to 6.4% (3+). Other changes with difficulty of reaching include a 3.5, 2.5, and 1.0 point change of the difference between, respectively, those with and without children in the household, those with and without a college degree (baseline: only high school), and men versus women.

These differences in panels A and B between the difficult and the easy to reach—and hence potentially between nonrespondents and respondents—cannot be eliminated by standard reweighing schemes. The generalizability of in-sample estimates to population-wide estimates regarding labor force participation therefore depends crucially on maintaining low nonresponse rate in the CPS (while managing high response quality, including accurate responses to individual questions).

Unemployment rate

In Table 3.3, which is otherwise identical to Table 3.2, we switch to analyzing the unemployment rate by replacing the dependent variable with a 0/1 unemployment indicator and limiting the sample to labor force participants. Beginning again with the adjusted means in panel B, we see a distinctive trend as the unemployment

rate drops from 8.0% (1 attempt) to 7.5% (2 attempts) to 6.5% (3+ attempts). This overall drop of 1.5 percentage points is similar to the drops in the annual unemployment rate from its Great Recession peak (9.6% in 2010) to its level in 2012 (8.1%) and from its 2001-recession peak (6.0% in 2003) to the next trough (4.6% in 2006–2007). It is again clear that unless nonrespondents are selected at random—a suspect assumption, given that even among respondents, difficulty of reaching is a strong predictor of the outcome—high response rates (with high-quality responses) are crucial for producing accurate population-wide estimates.

Looking at interactions in panel A, there appears to be less movement in Table 3.3 than in Table 3.2. (The only consistent trends with some statistical significance—up to the 5% level—are within some of the age rows.)

Table 3.3: Unemployment rate

Attempts	1	2	3+	NR
	A: Regression with interactions			
	Base	Interactions		
Age: 16–19	0.079***	0.012	0.022	-0.011
	(0.007)	(0.014)	(0.017)	(0.020)
20–39	0.014***	0.001	-0.011**	-0.004
	(0.002)	(0.004)	(0.005)	(0.006)
50–64	0.003	-0.005	-0.010**	0.001
	(0.002)	(0.004)	(0.005)	(0.006)
65 and up	0.002	-0.001	-0.008	0.014
	(0.003)	(0.007)	(0.008)	(0.009)
Children in household	-0.002	-0.009*	-0.006	-0.007
	(0.002)	(0.005)	(0.006)	(0.007)
Female	-0.004***	-0.005	-0.001	0.004
	(0.002)	(0.003)	(0.004)	(0.004)
Educ: Less than high school	0.043***	-0.001	-0.011	-0.002
	(0.004)	(0.008)	(0.009)	(0.013)
Some college or tech. school	-0.020***	0.003	0.000	0.003
	(0.002)	(0.004)	(0.005)	(0.006)
College graduate	-0.043***	0.003	0.008*	0.007
	(0.002)	(0.004)	(0.005)	(0.005)
Race: Black	0.064***	-0.018**	-0.010	-0.009
	(0.004)	(0.007)	(0.008)	(0.010)
Asian	-0.007*	0.006	0.010	-0.023***
	(0.004)	(0.007)	(0.008)	(0.008)
Other	0.047***	-0.012	0.003	-0.000
	(0.006)	(0.012)	(0.015)	(0.017)
Constant	0.061***	-0.032*	-0.025	-0.037
	(0.010)	(0.019)	(0.018)	(0.027)
	B: Adjusted means			
Unemployment rate	0.080***	0.075***	0.065***	0.066***
	(0.001)	(0.001)	(0.002)	(0.002)

Notes: Source: Current Population Survey, Jan. 2012–Dec. 2013. Sample: labor force participants, $N = 200,358$ (1 attempt: 124,530; 2: 35,376; 3+: 23,705; None Reported: 16,747). $R^2 = 0.05$. Table reports estimates from a single OLS regression (see page 85 for full explanation of table structure). Dependent variable: 0/1 unemployed indcator. Panel A: estimated coefficients from a fully interacted regression: each regressor is interacted with each difficulty-to-reach category (omitted category: 1 attempt). Regression also includes non-reported indicators (and their interactions) for marital status (6 categories), household size (5), state (51), urban/rural (3), interview month (12), and interview year (2); see appendix note A.1. Standard errors, clustered at the household level, in parentheses. Panel B: adjusted means, calculated from panel A regression. *** p<0.01, ** p<0.05, * p<0.1.

3.2.3 Robustness and Additional Results

Self reports versus proxies

Unlike the CEX—where a respondent provides expenditure information at the household level—and BRFSS—where a respondent provides individual information—in the CPS a respondent provides information about herself *and* about every other adult in the household. This means that in the CPS, number of contact attempts may be a noisier measure of the true, underlying difficulty of reaching household members who are not personally interviewed. Are the difficulty-of-reaching trends we report above indeed stronger among self reporters?[10]

To explore this hypothesis, appendix Tables B.2 and B.3 (participation) and B.4 and B.5 (unemployment) reproduce Tables 3.2 and 3.3 separately for self and proxy reports. As hypothesized, the difficulty-outcome trends are stronger—indeed, much stronger—among self reporters. Adjusted means for labor force participation show a monotonic 7.5-point increase (from 62.4 to 69.9 points) among self reporters, compared with an only 2.3-point increase (from 65.8 to 68.1) among the proxy reported. There are also generally more pronounced difficulty-of-reaching trends within the demographic-group estimates of self reporters. Similarly, when moving from 1 to 2 to 3+ attempts, adjusted means for unemployment display a monotonic 2.5-point decrease (8.4% to 7.2% to 5.9%) among self reporters, again roughly three times the 0.7-point decrease among the proxy reported (7.7% to 7.8% to 7.0%, no longer strictly monotonic).

[10]Notice that one should not expect the difficulty-of-reaching trends to entirely disappear among proxy reports, for reasons that could be both mechanical—e.g., all else equal, the household of a harder-to-reach individual is, on average, harder to reach—and circumstantial—e.g., demographic predictors of difficulty-of-reaching (such as age) are correlated across members within a household.

While the dramatically steeper trends among self reporters in the adjusted means of both dependent variables support the idea that there is a link between the difficulty of contacting a respondent and her outcomes, and that the link cannot be accounted for with other observable demographic controls, the comparison of self versus proxy reports should be interpreted with caution, as it has its own set of limitations. Importantly, self reporters are likely to be different from those whose labor force status is reported by proxy. Indeed in our data self reporters are on average older, more highly educated, and more likely to be female (not reported); of course, they may also be different on unobservable characteristics. We therefore refrain from drawing strong conclusions based on self reporters alone. (Our main analysis above pools together proxy- and self-reports, more closely matching the sample used by the BLS when it estimates its statistics.)

Telephone-completed interviews only

As mentioned above, the CPS's number-of-contact-attempts variable only counts in-person attempts. In addition, the interview-mode variable only records the mode of the last interview (in person, telephone, or missing).[11] One may therefore wonder whether a respondent whose last interview was by telephone may on average be harder to reach than a respondent whose last interview was in person, given the same number of reported in-person contact attempts, as the former required at least one (and possibly more than one) additional, unrecorded, telephone contact attempt.[12] Specifically, consider the hypothesis that at a given number of reported

[11] We learned in conversations with BLS staff that the mode variable gets set every time the interviewer enters the case. As a result, for households whose data were collected over the course of multiple interviews, only the last interview's mode is recorded.

[12] Of course, as we do not observe the total number of telephone contact attempts, we can neither quantify the hypothesized difficulty difference nor can we even be certain that it exists, as unrecorded telephone attempts could have taken place also for individuals who completed the interview in person.

in-person attempts, telephone completes are more likely to have had, on average, an additional unobserved positive number of telephone attempts. This hypothesis yields two predictions: relative to in-person completes, the 17.1% telephone completes in our main sample should on average (a) have outcomes consistent with being more difficult than their difficulty measure suggests, and (b) show weaker outcome-difficulty trends (due to the difficulty variable being a noisier measure for them).

Appendix Tables B.6 (participation) and B.7 (unemployment) repeat our main analysis restricting the sample to the 52,624 telephone completes in our data. (We do not report estimates restricted to in-person completes because, accounting for 80.8% of our data, they are rather close to the full-sample estimates above.) The adjusted means for labor force participation follow the same upward trend as in Table 3.2, rising from 68.5% to 70.3% to 71.3% across the difficulty-to-reach categories, but are generally higher than the full-sample adjusted means (64.1%, 66.2% and 69.0%), consistent with (a) above, and their trend is less steep, consistent with (b) above. Likewise, the adjusted means for unemployment follow the same pattern as in Table 3.3, declining from 6.4% to 6.0% to 5.1%, are lower than their full sample counterpart (8.0%, 7.5% and 6.5%, respectively), and show a slightly less steep trend although not statistically significantly so. The coefficient patterns in panel A of both appendix tables generally resemble those in the primary analysis, but at different levels (similar to the findings in the adjusted means). Overall, then, the evidence is consistent with the above hypothesis.

3.3 BRFSS: Obesity

3.3.1 Data

The Behavioral Risk Factor Surveillance System (BRFSS) is a large annual cross-sectional telephone survey designed to monitor health-related risk behaviors, chronic health conditions, and the use of preventative health services among the adult U.S. population. Most modules of the survey are designed and supported by the Centers for Disease Control and Prevention (CDC). However, data collection is decentralized and is individually administered by each of the 50 states, the District of Columbia, Guam and Puerto Rico (henceforth, the 53 locations). The survey is intended to be nationally representative of adults living in households in the United States and these territories. It uses random digit dialing for both landline and cellular telephone numbers; for landlines, one respondent is randomly chosen per household. The survey is conducted throughout the year, with differences in the exact timing between the 53 locations.[13] The 2012 BRFSS dataset—the latest publicly available at the time of conducting our analysis—includes 475,687 respondents. Across the 53 locations, its median survey response rate is 45.2%, ranging from 27.7% in California to 60.4% in South Dakota; weighted by location sample size, its average response rate is 44.6%.[14]

The BRFSS contains a variable with respondents' Body Mass Index (BMI,

[13]Appendix Figures B.1 through B.3, each containing 54 mini-graphs representing the entire sample and the 53 locations, report the corresponding 54 distributions of contact attempts and of interview timing information (the month, day of the month, and day of the week in which the interview was conducted).

[14]The BRFSS calculates response rates using AAPOR Response Rate #4, separately for its landline and cellular phone samples within each location; it then reports location response rates as weighted averages, weighted by local landline/cellular full-sample sizes. See BRFSS (2013) for a detailed discussion of response rate calculation and variation across locations.

defined as $\frac{\text{mass in kg}}{(\text{height in meters})^2}$). It is calculated from respondents' reports of their height and weight (in either feet and pounds or meters and kilograms), and is missing for respondents who did not provide their height or weight ($n = 22,710$, 4.8% of the original sample); who reported being pregnant at the time of the interview ($n = 2,680$, 0.6%); or whose calculated BMI was considered erroneous by the CDC ($n = 85$). Our analysis is based on the remaining 450,212 respondents. Our main dependent variable is a 0/1 indicator of obesity, defined by the World Health Organization as BMI ≥ 30 (WHO, 1995, 2000).

Our difficulty-of-reaching measure is a variable that records the number of calls made to each respondent. We use it to divide the data into four approximate difficulty quartiles: 1 call (25.5% of the data), 2–3 calls (30.5%), 4–6 calls (21.9%) and 7 or more calls (22.1%).[15]

For completeness and transparency, we note that chronologically, BRFSS data were the first we analyzed, and prior to analyzing obesity we analyzed an outcome that, while not a key outcome of our paper, was of great interest to us nonetheless: a life satisfaction question. We first attempted to replicate the main findings in Heffetz and Rabin (2013) and then, as BRFSS data have been used by economists to study the relationships between life satisfaction and properties of states (Oswald and Wu, 2010) and cities (Glaeser et al., 2016), we also explored the state relationships by difficulty of reaching and by cross-location differences in survey methodology. Our early analysis of the life satisfaction question used older (2005–2008) BRFSS data, matching the dataset in Oswald and Wu (2010); in general we replicated the published results that we explored.[16]

[15] In the public-use data we analyze, the number-of-calls variable is globally top-coded at 35, and in some states it appears to be locally top-coded at 15 (top-coding does not affect our difficulty categories). Appendix Figure B.4 reports the distribution of this variable in our sample and in each of the 53 locations.

[16] Specifically, Heffetz and Rabin (2013) found that cross-group differences in reported hap-

3.3.2 Analysis by Difficulty of Reaching

Sample composition

Table 3.4 reports the demographic composition of our BRFSS sample. As with the CPS, we find compositional differences in the makeup of the difficulty-to-reach subsamples. Notably, in both datasets the difficult to reach tend to be younger than the easy to reach. But in contrast with the CPS, in the BRFSS females are on average easier to reach: their share decreases monotonically from 61.2% (1-attempt respondents) to 56.3% (7+ attempts). The BRFSS sample also becomes less non-hispanic white (79.5% to 73.8%), more educated, and of higher reported income with difficulty of reaching. While the relevant variable definitions are not directly comparable across the CPS and BRFSS, these trends seem consistent across the two datasets. In particular, higher income is consistent with higher participation and lower unemployment (see also appendix Table B.1, which replicates table 1 and adds CPS income data). Finally, there is a strong, monotonic and significant (both economically and statistically) difficulty-of-reaching trend in the fraction of the sample that is obese, which declines from 29.4% (1 attempt) to 27.1% (7+ attempts). Interestingly, this trend does not reflect a trend in the average weight of the difficulty-to-reach categories (as weight remains relatively constant), though it may in part reflect a 1 cm overall increase in average height from the 1- to

piness depend on the difficulty of reaching respondents; our appendix Tables B.23 and B.24, modeled after tables 1–4 from Heffetz and Rabin (2013), report that the original results replicate qualitatively (the data show the same directional patterns), but the estimated magnitudes are smaller. While this may be seen as a successful replication in a new dataset, we caution that the original outcome variable (happiness yesterday) is not directly comparable to the BRFSS variable (general satisfaction with one's life). Oswald and Wu (2010) found that average life satisfaction in a U.S. state is related to non-subjective quality-of-life measures for the state; we replicated their analysis and found that controlling for cross-state differences in survey methodology—including the faction of a state's population that was easy-to-reach, the faction interviewed in each third of the month, and the state's overall response rate—did not affect the original findings.

7+-attempt categories.

Table 3.4: BRFSS Demographics

Attempts:		1	2–3	4–6	7+	All
Age:	18–39 (%)	18.5	20.6	23.6	20.9	20.8
		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
	40–49	12.1	13.6	15.5	17.6	14.5
		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
	50–59	18.3	19.8	21.1	23.7	20.6
		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
	60–69	22.1	21.4	20.0	20.5	21.1
		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
	70 and up	28.5	24.1	19.2	16.6	22.5
		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
	Missing	0.5	0.6	0.6	0.7	0.6
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Children in household		22.4	25.5	29.3	31.8	26.9
		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
Female		61.2	58.2	56.5	56.3	58.2
		(0.1)	(0.1)	(0.2)	(0.2)	(0.1)
Educ:	Less than high school	9.0	8.7	8.5	8.6	8.7
		(0.1)	(0.1)	(0.1)	(0.1)	(0.0)
	High school	30.7	29.6	28.7	28.2	29.4
		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
	Some college or tech. school	27.8	27.5	27.1	25.6	27.1
		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
	College graduate	32.5	33.9	35.5	37.4	34.7
		(0.1)	(0.1)	(0.2)	(0.2)	(0.1)
	Missing	0.1	0.2	0.2	0.2	0.2
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Race:	White, non-hispanic	79.5	78.0	75.2	73.8	76.8
		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
	Black, non-hispanic	7.5	7.8	8.5	10.1	8.4
		(0.1)	(0.1)	(0.1)	(0.1)	(0.0)
	Other, non-hispanic	6.0	6.4	6.7	6.4	6.4
		(0.1)	(0.1)	(0.1)	(0.1)	(0.0)
	Hispanic	5.9	6.7	8.3	8.4	7.3
		(0.1)	(0.1)	(0.1)	(0.1)	(0.0)
	Missing	1.0	1.1	1.2	1.3	1.2
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Inc:	Below $25,000	30.2	27.3	25.6	22.8	26.7
		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
	$25,000–49,999	23.9	23.4	22.7	21.3	22.9
		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
	$50,000–74,999	12.9	13.7	14.0	14.2	13.7
		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
	$75,000 and up	20.0	22.7	25.2	29.2	24.0
		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
	Missing	13.0	12.8	12.5	12.5	12.7
		(0.1)	(0.1)	(0.1)	(0.1)	(0.0)
Obese ($BMI \geq 30$)		29.4	28.5	28.1	27.1	28.4
		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
Avg. height (cm)		168.7	169.3	169.6	169.7	169.3
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Avg. weight (kg)		79.6	79.8	79.9	79.7	79.8
		(0.1)	(0.1)	(0.1)	(0.1)	(0.0)
Median number of attempts		1	2	5	10	3
Observations		114,694	137,418	98,813	99,287	450,212

Notes: Source: Behavioral Risk Factor Surveillance System, 2012. Sample: Non-pregnant individuals who provided height and weight (and were not excluded due to erroneously high/low BMI). All figures (and standard errors) reflect proportions within each column's difficulty-of-reaching category, except for those for height and weight (which report averages) and number of attempts (which report medians).

Obesity

Table 3.5 follows the same structure as the CPS outcome tables (see table 3.2's exposition on p. 85). It reports the results from a regression of a 0/1 obesity indicator on a set of demographic indicators (those in Table 3.4 and additional unreported indicators for marital status (7 categories), location (53), urban/rural (6), and interview month (12)) and a constant, a set of difficulty-of-reaching indicators, and a full set of demographic-difficulty interactions. As panel B reports, adjusted obesity prevalence declines monotonically from 29.7% (1 attempt) to 26.6% (7+ attempts), an overall decrease of 3.1 percentage points that is in fact larger than the unadjusted mean decrease of 2.3 points. Not only do differences in demographic composition across the difficulty-to-reach groups not drive the differences in obesity prevalence; the demographic-composition differences in fact *mask* some of the adjusted differences in obesity across difficulty categories. Since our entire sample could be viewed as the 44.6% easiest to reach among eligible households, the sample average of 28.4% obesity may be an overestimate of the population average. For example, a simple out-of-sample extrapolation of the within-sample trend in Table 3.5 suggests an overestimate of around 2 additional percentage points for the population obesity prevalence.[17] Of course, such a simple extrapolation leaves out many important details (including, for example, the possibility of measurement error that is correlated with difficulty), and is hence only given as an illustration. Our point is that our data strongly question the practice of assuming that nonresponders are on average as obese as responders (unconditionally or conditionally on observables), and accordingly regarding the sample average (raw or reweighted)

[17]This back-of-the-envelope estimate makes the simplifying assumptions that our four observed difficulty-to-reach categories are equal-sized and represent the easiest-to-reach half of the total attempted population. The remaining, unobserved, half of the population is then split into 4 equal-sized difficulty-to-reach groups; a linear trend that approximates the trend observed in panel B's adjusted means is projected onto these groups; and then the overall average is taken.

as the best population-wide estimate.

In panel A we note two cross-demographic-group obesity differences that change with difficulty of reaching and that we find of particular interest. First, women change from being essentially as obese as men (difference = −0.4 percentage points, insignificant) among 1-attempt respondents to being 2.2 points less obese than men among 7+-attempts respondents, a large and significant overall change of 1.8 points. Second, those with children in the household change from being a significant 2.3 points more obese among 1-attempt respondents to being essentially as obese as those in children-less households (difference = 0.5 points, insignificant), another large and significant overall change of 1.8 points. While neither of these changes is strictly monotonic (see table), in both cases conclusions regarding cross-demographic-group differences in obesity rates—comparing women versus men, and those with children versus without—qualitatively change. In both cases, whether one estimates a large and statistically significant difference or no difference depends on which difficulty category one looks at, and—given the BRFSS's relatively low response rate—population-wide inferences strongly depend on what one assumes about nonresponders.[18]

[18]Appendix Table B.8 reproduces Table 3.5 with additional controls for reported employment status. The table shows some interesting patterns: relative to those employed for wages—other things equal—homemakers, the self-employed, and especially students, are less obese; while those out of work for more than a year, and especially those unable to work, are more obese. Also, as expected, some of the education and income coefficients change. At the same time, both the adjusted means in panel B and the patterns noted above in panel A change little, and always remain within their original standard errors.

Table 3.5: Obesity

Attempts		1	2–3	4–6	7+
		A: Regression with interactions			
		Base	Interactions		
Age:	18–39	-0.093***	0.007	0.009	0.031***
		(0.005)	(0.007)	(0.007)	(0.007)
	40–49	-0.019***	0.013*	0.001	0.011
		(0.005)	(0.007)	(0.007)	(0.007)
	60–69	-0.007*	0.004	0.005	0.004
		(0.004)	(0.006)	(0.006)	(0.006)
	70 and up	-0.122***	0.012**	0.007	0.028***
		(0.004)	(0.006)	(0.007)	(0.007)
Children in household		0.023***	-0.007	-0.004	-0.018***
		(0.004)	(0.005)	(0.006)	(0.006)
Female		-0.004	-0.006*	-0.021***	-0.018***
		(0.003)	(0.004)	(0.004)	(0.004)
Educ:	Less than high school	0.025***	-0.000	0.001	0.006
		(0.005)	(0.007)	(0.008)	(0.008)
	Some college or tech. school	-0.007**	0.005	-0.002	0.005
		(0.003)	(0.005)	(0.005)	(0.005)
	College graduate	-0.064***	0.002	-0.006	-0.007
		(0.004)	(0.005)	(0.005)	(0.005)
Inc:	Below $25,000	0.035***	-0.003	-0.009	-0.016**
		(0.005)	(0.006)	(0.007)	(0.007)
	$25,000-49,999	0.009*	-0.005	-0.006	-0.004
		(0.005)	(0.006)	(0.007)	(0.007)
	$75,000 and up	-0.045***	0.003	0.001	0.008
		(0.005)	(0.006)	(0.007)	(0.007)
Race:	Black, non-hispanic	0.123***	0.000	-0.010	0.013*
		(0.005)	(0.007)	(0.008)	(0.007)
	Other, non-hispanic	-0.005	0.010	0.021**	-0.001
		(0.006)	(0.008)	(0.009)	(0.009)
	Hispanic	0.033***	0.001	0.003	0.003
		(0.007)	(0.009)	(0.009)	(0.009)
Constant		0.370***	0.013	0.038**	-0.009
		(0.012)	(0.016)	(0.017)	(0.018)
		B: Adjusted means			
Obesity		0.297***	0.287***	0.282***	0.266***
		(0.001)	(0.001)	(0.001)	(0.002)

Notes: Source: Behavioral Risk Factor Surveillance System, 2012. $N = 450,212$ (1 attempt: 114,694; 2–3: 137,418; 4–6: 98,813; 7+: 99,287). $R^2 = 0.04$. Table reports estimates from a single OLS regression. Dependent variable: 0/1 obesity indicator. See page 85 for full explanation of table structure. Panel A: estimated coefficients from a fully interacted regression: each regressor is interacted with each difficulty-to-reach category (omitted category: 1 attempt). Regression also includes non-reported indicators (and their interactions) for missing data, marital status (7 categories), location (53), urban/rural (6), and interview month (12); see appendix note A.1. Standard errors in parentheses. Panel B: adjusted means, calculated from panel A regression. *** p<0.01, ** p<0.05, * p<0.1.

3.3.3 Robustness and Additional Results

Height and weight

The variables that underlie the obesity variable—self-reported weight (kg) and height (m)—may themselves be interesting to researchers as main outcomes. We repeat our analysis with each of the two as a dependent variable (appendix Tables B.9 and B.10). We find a strong decreasing trend in adjusted mean weight, which drops from 80.4 [SE 0.06] to 79.1 [0.06] kg from the easiest to hardest category—a 1.3 kg ($\approx$ 2.9 lbs) difference. We also find evidence of cross-group differences—including across males and females and across some age and income groups—with overall changes in the range 0.8–1.6 [0.2–0.3] kg from easiest to hardest.

In contrast, adjusted mean height remains remarkably stable, at 169.3 [0.02–0.03] cm, across the four difficulty-to-reach categories, with small ($<$ 1 [0.1] cm) changes from easy to difficult in cross-sex and cross-race differences (Table B.10).

Weighted regressions

Whereas the CPS and CEX are both multiple-interview panels from which we only analyze a single interview per household—MIS 1 in the CPS and second interview in the CEX (see next section)—the BRFSS is a cross section of individuals. As a result, unlike with the other two datasets, with the BRFSS we can use the provided (full sample) survey weights to probe the robustness of our unweighted analysis.

The BRFSS includes analysis weights designed to adjust for nonresponse and non-coverage in the survey and to make the number of cases sum to the population for each geographic region (usually state) (CDC, 2012). The weighting process

takes into account the probability that someone was likely to be sampled. The weight is also raked on up to 12 demographic margins.[19] Appendix Tables B.11, B.12, and B.13 recreate Tables 3.5 (obesity), B.9 (weight), and B.10 (height), respectively, using the provided individual weights. In the adjusted means for obesity, the weighted prevalence trends from 29.4% to 28.4% to 27.2% to 26.1% [all SEs 0.3] across increasing difficulty-to-reach categories (compared with 29.7% to 28.7% to 28.2% to 26.6% in the unweighted Table 3.5). We also find similar patterns in the adjusted means for weight and height across the weighted and unweighted tables, and for all three outcomes we additionally find the same general patterns, or lack thereof, in panel A's coefficients.

In summary, using the BRFSS weights does not change our in-sample conclusions.[20] This finding questions the assumption that BRFSS nonrespondents are similar to the (properly weighted) average respondent.

[19] These are: age group by gender, race/ethnicity, education, marital status, home ownership status, gender by race and ethnicity, age group by race and ethnicity, phone ownership, region, region by age group, region by gender, region by race and ethnicity. See CDC (2012) for additional details.

[20] As an additional robustness check, appendix Table B.14 recreates Table 3.5 based on the 303,034 respondents (67.3% of our sample) who were interviewed by one of the 1,651 interviewers who conducted at least 10 interviews in each difficulty category; and appendix Table B.15 does the same and adds a full set of interviewer controls and interviewer-difficulty interactions. As the tables show, our findings do not change, suggesting that using weights that accounted for such information would not change our conclusions. (We conduct this test only for BRFSS data because we do not observe interviewer ID in the CPS and CEX.)

3.4 CEX: Quarterly Household Expenditures

3.4.1 Data

The Consumer Expenditure Survey (CEX), administered by the Bureau of Labor Statistics, measures the purchasing habits of U.S. consumers. The CEX's random sample of households is designed to be representative of the non-institutionalized, civilian population of the U.S.[21]

The CEX consists of two separate components—a diary survey and an interview survey—with two independent samples. We focus on the interview survey, which collects data on up to 95% of household expenditures. Its goal is to collect detailed data on large purchases that respondents can be expected to remember for three or more months (accounting for 60–70% of household purchases), and estimates for other major categories of expenses (accounting for 20–25% of purchases). The survey is a rotating panel: each quarter, approximately 20% of the sample is new; each household participates for five consecutive quarters.

The first interview collects demographic and other household details. Each of the second through fifth interviews collects expenditure data for the three preceding months; these data are used by the BLS to revise the weights in the Consumer Price Index, as well as to create national expenditure estimates. The survey has a quarterly target of approximately 7,000 participating households. Response rates among eligible households for 2008 to 2013 were, respectively, 73.8%, 74.5%, 73.4%, 70.4%, 69.5% and 66.7%—a fast decline of roughly 7 percentage points in five

[21] This section is based on the documentation available at `http://www.bls.gov/cex/` (accessed on May 26, 2016). This includes the percentages of household expenditures that the CEX is estimated to cover (reported in the next paragraph).

years.[22]

Beginning in 2008, the BLS releases detailed paradata for the CEX interview component. The paradata contain information about each contact attempt, including when it occurred, its mode (in person or telephone), and its result. We create a number-of-contact-attempts variable by counting the contact-attempt entries in the paradata file. For consistency with our analyses in the previous sections, in our main analysis we use this variable for the first interview. However, since no relevant expenditure data are collected in the first interview, we use the second interview's expenditure data. (When we also use the second interview's number of contacts, our results below become stronger; see footnote 23.)

Our primary estimation sample is 2008:Q2–2013:Q4; we omit 2008:Q1 because we do not have paradata for its first interview, which occurred in 2007. We start with 39,277 observations, from which we drop 2,553 (6.5%) that could not be matched with first-interview paradata. We further exclude 2 observations from the main analysis of log total expenditures and 156 observations from the robustness analysis of log health expenditures due to negative expenditure values. We divide the resulting sample into four difficulty-of-reaching groups, as equal-sized as the contact-attempts distribution allows: 1 attempt (18.0% of the sample), 2 (19.7%), 3–4 (27.2%), and 5 or more (35.2%).

Our key outcome variable is log total quarterly expenditures (specifically, $ln[1+$ expenditures]). When reporting and discussing results, we often transform the estimates back to dollar values (by exponentiating and subtracting 1); we use the delta method to transform SEs.

[22] As with the CPS, these figures correspond with AAPOR's Response Rate #2, #4, and #6. (There are no households of unknown eligibility, as CEX interviewers classify the eligibility of each address.)

3.4.2 Analysis by Difficulty of Reaching

Sample composition

Table 3.6 reports substantial demographic differences across difficulty-to-reach groups in the CEX—a finding similar to those from the other two datasets. As in the CPS and BRFSS, the CEX's hard-to-reach are younger: the youngest two age categories, together covering ages 16–39, increase their share from 27.0% of the 1-attempt subsample to 33.5% of the 5+ subsample, while the 65+ age category falls from 30.9% to 15.6%. Also consistent with the other two datasets, the 5+ subsample is more educated and earns more than the 1-attempt subsample, with the proportions of college degree holders and of household income above $70,000 respectively increasing by 5.4 and 10.4 percentage points.

There is also a qualitatively large and statistically significant trend in our key outcome, log total expenditures: exponentiated, they increase from $8,225 (1 attempt) to $9,200 (2), to $9,865 (3–4), to $9,990 (5+). Finally, the table also reports averages for two arguably important expenditure categories that have received much attention from economists and that we chose ahead of time (prior to looking at the data for specific categories): food and health. Average log food expenditures mirror the total expenditures pattern of monotonic (and highly statistically significant) increase, while health expenditures appear to generally (though nonmonotonically) decrease.

Table 3.6: CEX Demographics

Attempts		1	2	3–4	5+	All
Age:	16–29 (%)	13.2	10.0	11.0	13.5	12.1
		(0.4)	(0.4)	(0.3)	(0.3)	(0.2)
	30–39	13.8	15.2	17.1	20.0	17.2
		(0.4)	(0.4)	(0.4)	(0.4)	(0.2)
	40–49	14.3	18.7	19.8	22.1	19.4
		(0.4)	(0.5)	(0.4)	(0.4)	(0.2)
	50–64	27.8	28.9	30.2	28.8	29.0
		(0.6)	(0.5)	(0.5)	(0.4)	(0.2)
	65 and up	30.9	27.2	22.0	15.6	22.4
		(0.6)	(0.5)	(0.4)	(0.3)	(0.2)
Children in household		28.4	32.0	33.4	36.3	33.3
		(0.6)	(0.5)	(0.5)	(0.4)	(0.2)
Female		53.6	53.3	52.8	53.2	53.2
		(0.6)	(0.6)	(0.5)	(0.4)	(0.3)
Educ:	Less than high school	15.9	14.7	13.2	12.2	13.7
		(0.5)	(0.4)	(0.3)	(0.3)	(0.2)
	High school or GED	26.5	25.8	24.6	24.2	25.1
		(0.5)	(0.5)	(0.4)	(0.4)	(0.2)
	Some college or tech. school	30.4	28.8	29.9	30.9	30.1
		(0.6)	(0.5)	(0.5)	(0.4)	(0.2)
	College graduate	27.2	30.7	32.3	32.6	31.2
		(0.5)	(0.5)	(0.5)	(0.4)	(0.2)
Race:	White	84.3	83.0	82.3	77.8	81.2
		(0.4)	(0.4)	(0.4)	(0.4)	(0.2)
	Black	10.2	10.3	11.2	14.2	11.9
		(0.4)	(0.4)	(0.3)	(0.3)	(0.2)
	Asian	3.9	4.7	4.7	5.5	4.8
		(0.2)	(0.2)	(0.2)	(0.2)	(0.1)
	Other	1.7	2.0	1.9	2.5	2.1
		(0.2)	(0.2)	(0.1)	(0.1)	(0.1)
Inc:	Below $20,000	26.4	21.4	19.1	18.3	20.6
		(0.5)	(0.5)	(0.4)	(0.3)	(0.2)
	$20,000-39,999	24.8	23.0	21.3	21.1	22.2
		(0.5)	(0.5)	(0.4)	(0.4)	(0.2)
	$40,000-69,999	22.4	24.0	24.1	23.7	23.6
		(0.5)	(0.5)	(0.4)	(0.4)	(0.2)
	$70,000 and up	26.4	31.7	35.5	36.8	33.6
		(0.5)	(0.5)	(0.5)	(0.4)	(0.2)
Exp:	Total (log)	9.02	9.13	9.20	9.21	9.15
		(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
	Total ($)	8,225	9,200	9,865	9,990	9,459
		(74)	(79)	(72)	(64)	(36)
	Health (log)	5.12	5.20	5.12	4.83	5.03
		(0.03)	(0.03)	(0.03)	(0.03)	(0.01)
	Food (log)	7.15	7.24	7.31	7.33	7.27
		(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Median number of attempts		1	2	3	7	3
Observations		6,603	7,230	9,973	12,918	36,724

Notes: Source: Consumer Expenditure Survey, 2008–2013. Sample: all second interview households that could be matched with a first-interview difficulty measure, excluding 2 and 156 observations with negative entries, respectively, in the total expenditure rows and in the health expenditure row. All figures (and standard errors) reflect proportions within each column's difficulty-of-reaching category, except for those for expenditures (which report average log expenditures and average log expenditures exponentiated into dollars), and number of attempts (which report medians).

Total expenditures

Table 3.7 has the same format as previous main-results tables. Panel A presents regression results for households' log quarterly total expenditures. Panel B's adjusted means are calculated from the panel A regression, and are reported both directly (in logs) and transformed back into dollars. The latter increase from $9,120 (amongst the 1-attempt subsample) to $9,331 (2), to $9,581 (3–4), and then stay flat at $9,585 (5+). Thus, accounting for the changing demographic composition of the subsamples, the difference between the easiest- and hardest-to-reach respondents shrinks from $1,765 in the unadjusted means (Table 3.6) to $465—a still very significant 5% increase—in the adjusted means (Table 3.7).[23] We find no significant trends in the coefficient estimates in Panel A.

[23] We view these estimates as somewhat conservative. Recall that we pair a household's difficulty-to-reach measure (from the first interview) with its outcome variable (expenditures from the second interview) across two different quarters. Therefore, if a household's difficulty of reaching in a specific interview is more strongly related to its outcome measures in that same interview, our estimates might be attenuated. To explore this possibility, appendix Table B.16 replicates Table 3.7, keeping the sample constant but using the difficulty-to-reach measure from each household's second interview. We indeed find similar patterns overall, with a larger expenditure gap between easiest- and hardest-to-reach respondents: the difference increases from $465 (Table 3.7) to $769 (Table B.16), reflecting both a drop from $9,120 to $8,942 among the easiest and a slightly smaller increase, from $9,585 to $9,711, among the hardest to reach. Furthermore, including additional controls for interview hour and duration—which are not available in our CPS and BRFSS data, and hence are not included in our main specification, but could in principle be incorporated into the survey weights—only decreases the (now nonmonotonic) expenditure gap in table 3.7 to $379 (Table B.17)—still a very significant 4% increase—and the expenditure gap in Table B.16 to $427 (Table B.18).

Table 3.7: Total expenditures

Attempts		1	2	3–4	5+
			A: Regression with interactions		
		Base	Interactions		
Age:	16–29	-0.091***	-0.010	-0.004	0.012
		(0.025)	(0.035)	(0.032)	(0.030)
	30–39	-0.025	0.002	0.000	-0.022
		(0.023)	(0.030)	(0.028)	(0.027)
	50–64	0.000	-0.030	-0.002	-0.021
		(0.021)	(0.028)	(0.026)	(0.024)
	65 and up	-0.026	-0.029	-0.020	-0.046
		(0.023)	(0.031)	(0.029)	(0.028)
Children in household		0.063***	-0.048	-0.041	-0.022
		(0.023)	(0.031)	(0.029)	(0.027)
Female		-0.022*	0.019	0.006	0.025*
		(0.012)	(0.017)	(0.016)	(0.015)
Educ:	Less than high school	-0.152***	0.032	0.031	0.011
		(0.019)	(0.027)	(0.026)	(0.025)
	Some college or tech. school	0.104***	-0.018	-0.017	-0.031
		(0.016)	(0.023)	(0.021)	(0.020)
	College graduate	0.256***	-0.004	-0.013	-0.021
		(0.017)	(0.024)	(0.022)	(0.021)
Race:	Black	-0.076***	-0.019	-0.042	-0.016
		(0.020)	(0.028)	(0.026)	(0.024)
	Asian	-0.076**	-0.008	-0.030	0.024
		(0.032)	(0.042)	(0.039)	(0.037)
	Other	0.058	-0.125**	-0.115*	-0.037
		(0.047)	(0.063)	(0.059)	(0.055)
Inc:	Below $20,000	-0.614***	-0.020	0.008	0.034
		(0.019)	(0.027)	(0.025)	(0.024)
	$20,000-39,999	-0.257***	-0.011	0.014	0.025
		(0.018)	(0.025)	(0.023)	(0.022)
	$70,000 and up	0.396***	0.017	0.023	0.035
		(0.018)	(0.024)	(0.022)	(0.022)
Constant		9.172***	0.074	0.114**	0.022
		(0.037)	(0.051)	(0.048)	(0.046)
			B: Adjusted means		
Total expenditures (log)		9.118***	9.141***	9.168***	9.168***
		(0.006)	(0.006)	(0.005)	(0.004)
Total expenditures ($)		9,120***	9,331***	9,581***	9,585***
		(59)	(54)	(47)	(43)

Notes: Source: Consumer Expenditure Survey, 2008–2013. Sample: All households from Table 3.6, excluding 2 households with negative net total quarterly expenditures. $N = 36,722$ (1 attempt: 6,603; 2: 7,229; 3–4: 9,972; 5+: 12,918). $R^2 = 0.56$. Table reports estimates from a single OLS regression. Dependent variable: ln(total quarterly household expenditures +1). See page 85 for full explanation of table structure. Panel A: estimated coefficients from a fully interacted regression: each regressor is interacted with each difficulty-to-reach category (omitted category: 1 attempt). Regression also includes non-reported indicators (and their interactions) for marital status (5 categories), size of consumer unit (3), urban/rural (2), interview month (12), and interview year (6); see appendix note A.1. Standard errors in parentheses. Panel B: adjusted means, calculated from panel A regression. *** p<0.01, ** p<0.05, * p<0.1.

3.4.3 Robustness and Additional Results

Health and food expenditures

Appendix Tables B.19 and B.20 replicate Table 3.7, with log health and log food expenditures as dependent variables. As with the unadjusted means, the trend in food expenditures is in line with the trend in total expenditures. It shows a 6% increase from easiest- to hardest-to-reach respondents—close to the 5% increase in total expenditures. In contrast, health expenditures show a highly significant (both statistically and economically) 13% *decrease* from easiest to hardest—opposite in sign and over double the percent increase in total (and in food) expenditures. As in Table 3.7, there are no clear trends (and few significant interaction estimates) in panel A in either Table B.19 or B.20.

This finding of significant and opposite trends in the adjusted means for food and health—the only two consumption categories we looked at—suggests that while some expenditure categories may be underestimated in the entire population, others may be *over*estimated. It highlights that without empirical evidence, it is difficult to know a priori the direction of potential bias.[24]

Telephone and in-person interviews

As with the CPS, CEX interviews can occur either in person or over the telephone. In contrast with the CPS, in which only in-person attempts are recorded, in the

[24] A posteriori, this and the finding of no clear trends in cross-demographic-group differences in either food or health may be consistent with some speculative interpretations of our results. For example, these findings are consistent with the untested idea that within the demographic groups we linearly control for (age, income, having children, etc.), households with healthier lives and lifestyles (spending more on food and less on health) are on average more difficult to reach. However, see footnote 26 for an alternative interpretation.

CEX *all* contact attempt are supposed to be recorded, and we have no a priori reason to expect the number-of-contact-attempts variable to be a better or worse measure of difficulty across one mode or the other. Appendix Tables B.21 and B.22 replicate Table 3.7 separately for telephone and in-person interviews.[25] We observe the same general pattern in the in-person sample (24,305 interviews) and telephone sample (11,719 interviews) as we observe in the combined sample, but the in-person exponentiated adjusted means seem a few hundred dollars lower on average: $8,923 [SE $66] to $9,387 [54] (in person, a 5% increase) versus $9,344 [124] to $9,920 [71] (telephone, a 6% increase) from easiest to hardest.

3.5 Previous Related Work

3.5.1 Research Utilizing Difficulty-of-reaching Measures

Our paper contributes to an empirical literature that seeks to shed light on nonresponse bias by investigating the links between difficulty measures and survey-based outcomes. Within economics, we are aware of only two such investigations, both recent: Heffetz and Rabin (2013), discussed above; and Behaghel et al. (2014), discussed below. Outside of economics the literature is longer-established, larger, and growing. Heffetz and Rabin (2013) review it in some detail. This section focuses on new work that postdates their review, but we present it in the context of the earlier literature, organized by three strands: theoretically focused, empirically focused, and experimental.

[25] We use a variable indicating interview mode as reported by the interviewer. Interviews reported as occurring through mixed modes (telephone and in-person) and interviews with missing mode data are excluded.

First, on the theory-focused front, models have been developed where the probability of survey participation is related to the outcome of interest; see Potthoff et al. (1993), who also analyze published number-of-calls data to fit parameters in their model. Our findings in the present paper appear consistent with such models. A more recent literature builds on such earlier work and aims at further utilizing paradata, including number of contact attempts, to directly improve analysis. One idea is to incorporate paradata in reweighting methods. Biemer et al. (2013) present one such technique, a call-back model, in which the response propensity is modeled using information on the number of contact attempts. Krueger and West (2014) examine a similar but richer model that adds additional sources of auxiliary data, such as interviewer observations and characteristics of the interview location. Of particular interest to our analysis, the corrected weights sometimes move different subgroups' estimated outcome prevalence in opposite directions, suggesting that nonresponse bias may be differentially affecting different groups—consistent with our findings that the difficulty-outcome gradient may differ across demographic groups. Coming from economics rather than from statistics and survey methodology, Behaghel et al. (2014) (mentioned above) propose using difficulty-of-reaching paradata for dealing with survey attrition, and apply their proposed corrective to a job-search experiment.

Second, on the empirically focused front, we noted above two studies that link administrative data to survey data (see footnote 3). Such studies can shed light on both nonresponse bias—by comparing administrative-data outcomes across survey nonrespondents and respondents (of different difficulty)—and measurement error—by comparing respondent outcomes (by difficulty) across administrative and survey data. Lin and Schaeffer (1995) match child-support awards and payments with a telephone survey and examine nonresponse bias. They provide some evidence that

hard-to-reach respondents may not actually be more like nonrespondents than the easy-to-reach. However, their survey is conducted in a very different context than the large government surveys we study, as most of their nonrespondents were never located from the initial court records; and their resulting difficulty measure is also rather different, as they exclude contact attempts made prior to locating the respondent. Kreuter et al. (2010) link administrative and survey data for a subsample of German unemployment-benefits recipients who were interviewed by telephone, and provide insight on both nonresponse and measurement: in their data, adding hard-to-reach respondents reduces nonresponse bias (the true averages of survey participants become closer to the true target-population averages), but increases measurement error (survey reports of the hard to reach are farther from their true values). For three of the four outcomes they study—employment status, age, and foreign citizenship—the net effect is to decrease the overall bias in survey outcomes (the reduction in nonresponse bias outweighs the increase in measurement error). But for the receipt of unemployment benefits the addition of hard-to-reach respondents actually increases the overall bias in the survey measure. These findings highlight the concern that unwilling—and thus hard-to-reach—respondents may be more prone to misreport and, in particular, misreport sensitive behaviors.

Related to the potential unwillingness of the hard to reach, two recent papers propose respondents' motivation as a potential explanation of the link between difficulty and outcomes. In the first, a working paper, Chadi (2014) documents a connection between a respondent's motivation to participate in the survey, number of contact attempts, and subjective happiness. In the second, Meyer et al. (2015) examine the declining accuracy of survey data by combining survey and administrative measures of government transfers. They find that measurement error contributes as much to the observed bias in survey estimates as item nonresponse and

unit nonresponse combined; they speculate that an increase in "over-surveyed" and unmotivated participants could increase measurement error. While the idea that the hard-to-reach are less motivated and more prone to measurement error cannot be directly assessed in our data, it could not alone easily explain our findings. For example, to explain the cross-demographic-group trends we find for some outcomes, such difficulty-motivation links would have to differ across demographics. For another example, they would have to *increase* food expenditures and *decrease* health expenditures from easy to difficult respondents.[26] At the same time, the idea that for some variables, item nonresponse increases with difficulty of reaching finds support in our data. For example, based on reported imputation flags in the CPS, we calculate imputation rates for twelve-month family income to be 19.2 percent (1 attempt), 20.8 (2), 23.9 (3+), and 27.8 percent (NR). (There is no clear way to directly link imputation flags to any of our main-outcome variables.)

Third, on the experimental front, in an influential study Keeter et al. (2000) conduct two telephone surveys—one over five days with a response rate of 36%, the other over eight weeks with a response rate of 61%—and find mostly small differences in outcomes, with demographics a notable exception. While our evidence in the present paper is consistent with the finding of significant demographic-composition changes as response rates increase, we also find significant changes in all four (non-demographic) key outcome variables we examine. The latter suggests that the earlier findings should not be misinterpreted as providing a blanket

[26] Of course, the less motivated may be more reluctant, differentially across demographics and expenditures, to accurately report certain kinds of information. We cannot rule out, for example, that due to social-image considerations, the harder to reach over-report behaviors perceived as "positive"—labor force participation, total expenditures, and food expenditures—while under-reporting behaviors perceived as "negative"—unemployment, health expenditures, and (especially among women, all else equal) obesity. However, our finding (in Table 3.2) that the increase in labor force participation with difficulty of reaching is *less* pronounced (all else equal) among men, the childless, and college graduates does not easily fit within such a social-image explanation.

justification for drawing population-wide conclusions regarding non-demographic variables from surveys with low response rates.

While we are not aware of recent work investigating the generalizability of the above experimental findings by running more experiments, the experimental finding of demographic differences across easy- and hard-to-reach respondents has been highlighted in recent empirical work. That work acknowledges the possibility of a link between difficulty and outcome variables, but leaves open the possibility that demographic controls may alleviate the problem—something that we show in this paper is not generally possible. Legleye et al. (2013), for example, examine the effect of increasing the number of contact attempts in a French telephone survey designed to measure sexual and reproductive health (SRH) issues. The inclusion of harder-to-reach respondents, who are found to differ on SRH behaviors from the rest of the sample, is shown to make the sample closer to the demographic composition of the target population. Similarly, a working paper by Pudney and Watson (2013) simulates the impact of reducing the number of call attempts in two health and employment longitudinal surveys—the British Household Panel Survey (BHPS) and the Household, Income and Labour Dynamics in Australia (HILDA)—and find that it would change the samples' demographic composition and outcomes of interest such as disability, ill health and employment. Other recent papers also find significant differences in demographics and outcome variables between the easy and difficult to reach, including Cohen et al. (2013), and Hetschko and Chadi (2017).

Recall that we find a consistent demographic and socioeconomic gradient across our three datasets: hard-to-reach respondents are younger and more educated, and they have higher household income, than easy-to-reach respondents. In addition,

our outcome variables, which are associated with socioeconomic status, exhibit similar patterns: the difficult-to-reach are more likely to be employed (in the CPS) and are generally healthier (having lower obesity rates in BRFSS and lower medical expenditures in CEX). These findings are broadly consistent with those in the works reviewed above, although differences in methodology prevent direct comparisons.[27] This broad consistency highlights the fact that researchers interested in outcomes related to socioeconomic status may need to pay particular attention to the potential for nonresponse bias in their analyses.

3.5.2 Evidence on the Quality of Difficulty-of-reaching Measures

A limitation of paradata is that they are rarely collected for direct use by analysts. Often a mere by-product of the data collection process, their quality may be lower than that of other survey data. Bates et al. (2010) examine the quality of paradata across three federal surveys: the National Health Interview Survey (NHIS), the CEX, and the CPS. The three collect the data through a common Contact History Instrument (CHI).[28] The authors find that while most CHI entries in the CPS and NHIS were recorded immediately after the contact attempt, in the CEX almost 20% of the attempts were recorded with some delay. In all three datasets, attempts that did not result in a contact were more likely to be recorded later,

[27]For example, Legleye et al. (2013) and Pudney and Watson (2013) both find that the employed are harder to reach, while Curtin et al. (2000) find that those with higher income are harder to reach. Similarly, the existing evidence is consistent with the notion that the healthy are generally harder to reach than the unhealthy across a variety of measures such as annual medical expenditures, self-reported health, and disability status (Cohen et al., 2013; Pudney and Watson, 2013).

[28]The CHI information is publicly available—and is the source of our contact attempts data—for the CEX but not for the CPS.

and hence presumably had a higher chance of being forgotten, than those resulting in a contact. The authors also discuss an internal report according to which CEX interviewers estimated that they ever complete a record for only around 85% of their attempts. Similarly, Biemer et al. (2011) conduct an informal survey of field interviewers and supervisors for the National Survey on Drug Use and Health (NSDUH), and find greater incentives to underreport than to overreport the number of visit attempts.[29] To the extent that the uneven underreporting found in these studies adds measurement error to our difficulty-of-reaching independent variables, our estimates may be attenuated. Our estimated difficulty trends may hence be viewed as lower bounds.

3.6 Discussion and Conclusion

Investigating three of the most commonly used government surveys, we find significant, systematic differences in key outcomes across number-of-contacts groups. These differences persist within demographic cells—indeed, some of them systematically differ across demographic groups—and cannot be eliminated by standard reweighting. They qualitatively replicate within subsamples of the data, and they generally grow when the number-of-contacts measure seems cleaner.

That selection and nonresponse may bias survey outcomes is well-known theoretically but—judging by common practices—under-appreciated empirically. By demonstrating that even after adjusting for demographics, key outcome variables are strongly related, empirically, to respondents' difficulty of being reached—and

[29] According to the authors, overreporting could be caught through timesheet reviews while underreporting might help keep a case considered alive, thus helping the interviewer avoid being perceived as using time inefficiently.

hence, potentially, to their likelihood of survey participation—we hope to have convinced readers that a routine assumption of random nonresponse is hard to justify in these data.

In practice, how concerned should users of these and similar data be?

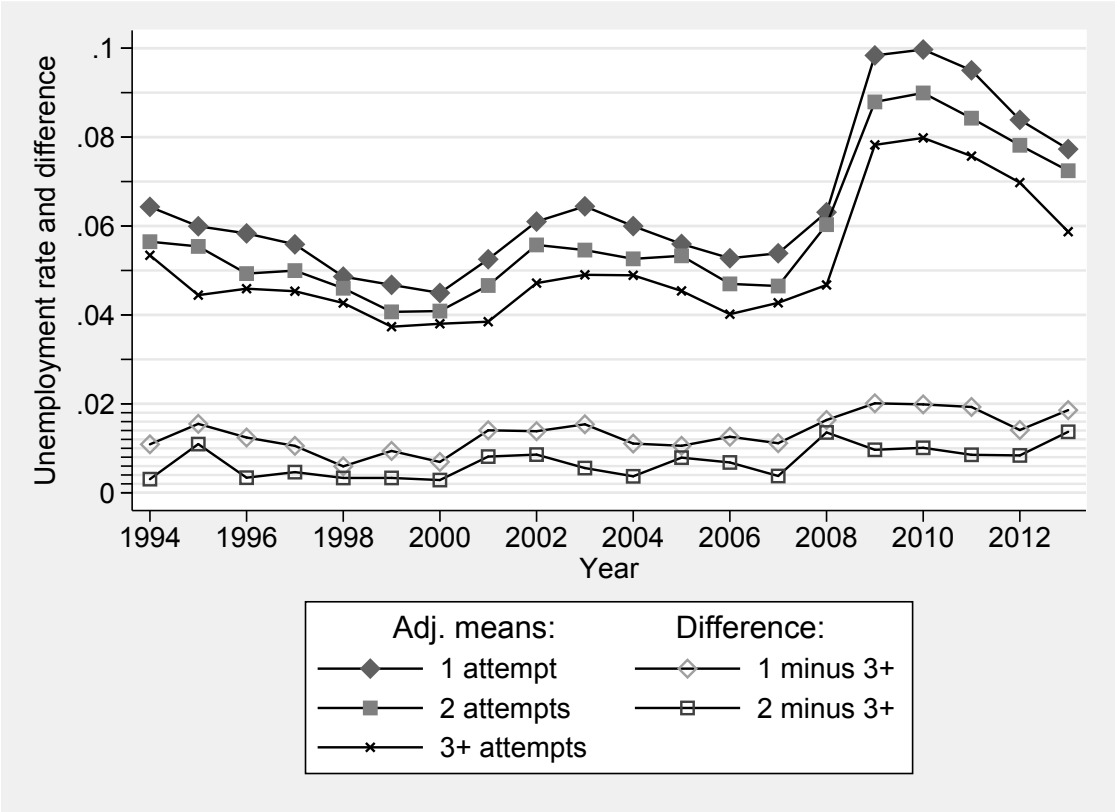
The answer depends on application and, importantly, on survey response rates: the lower these are, the more population-wide inferences from sample estimates rely on the assumption of random nonresponse (or random on observables). As we discussed when analyzing the BRFSS, our findings are consistent with the possibility that the population-wide prevalence of obesity could be 1 or 2 percentage points below the BRFSS sample mean, and, furthermore, cross-group comparisons of obesity (women versus men, for example) could change from an estimate of no difference to an estimate of a significant difference, depending on assumptions regarding nonresponders. Like the BRFSS, many telephone surveys nowadays have response rates below 50%—sometimes, well below that figure. In-person surveys still enjoy generally higher response rates, but they too show worrying trends: for example, as mentioned earlier, response rate in the CEX lost 7 percentage points from 2008 to 2013; this recent drop accelerated a previous, slower decline of 4 points from 2001 to 2008 (NRC, 2013). Brick and Williams (2012), who examine the causes of increasing nonresponse, note decreases in the response rate of several large, telephone and in-person, U.S. cross-sectional surveys from 1997–2007, including in the National Health Interview Survey (NHIS), National Household Education Survey (NHES), General Social Survey (GSS), and National Immunization Survey (NIS). Even among the most prominent in-person surveys, such as the CPS, response rates have been declining, dropping most recently from around 92% in 2010 to just above 89% in 2014 (Krueger et al., 2017). While the CPS's

presently high response rates mean that from a practical point of view, our CPS findings are mostly academic, and (measurement error aside) the official CPS-based population-wide estimates are likely rather accurate, our findings underline the importance of keeping response rates high—something that may or may not be achievable in the future.

We conclude with a concrete demonstration of the value of high (and accurate) response rates, in the context of one important application. Figure 3.1 tracks the unemployment rate by difficulty of reaching over time. The solid shapes depict the panel-B adjusted-mean estimates from twenty replications of Table 3.3, for the twenty years starting in 1994—the earliest year for which we have number-of-attempts data—with one exception: the "Children in household" control, which is not available for all years, is excluded from the regressions. The hollow shapes depict differences between these adjusted means. The figure generalizes our main finding from Table 3.3: in each of the twenty years, the unemployment rate is significantly higher among the easy than among the difficult to reach. However, the easy-minus-difficult difference (hollow squares) is both highly variable and, importantly, highly correlated with the unemployment rate (solid shapes).[30] In other words, unemployment fluctuations are more extreme among easy than among difficult respondents. Qualitatively, the figure therefore suggests that the common practice of treating nonrespondents as average respondents yields larger cyclical fluctuations in unemployment estimates relative to treating them as difficult respondents. Quantitatively, this striking finding is of negligible significance *at present* (given current CPS response rates)—but it demonstrates the importance

[30]The correlation between the 1-attempt unemployment rate (solid squares) and the 1-attempt-minus-3+-attempts difference (hollow squares) is 0.83 ($N = 20$). This high correlation is accompanied by substantial variation in the 1-attempt-minus-3+-attempts difference: for example, in 2000, when the U.S. annual unemployment rate was the lowest since the late 1960s, the difference was 0.7%; in 2010, when unemployment was the highest since the early 1980s, the difference was 1.9%.

Figure 3.1: Unemployment rate, by number of contact attempts and year



Notes: Source: Current Population Survey, 1994–2013. See details in text.

of keeping response rates and accuracy high.

While we know little about the increasingly many nonrespondents in the datasets we examine, we interpret our findings as, at the very least, suggesting that the burden of proof should lie with researchers whenever they make the random-nonresponse assumption in surveys with low response rates. Moreover, in specific applications this burden may extend across time and space. In particular, to the extent that the outcomes we examined are used for tracking the economy over time (as in the above example) or for making cross-country comparisons, our findings suggest that users should be concerned about, and try to control for, possible differences across periods and locations in the relation between outcomes and difficulty of reaching. Finally, this burden of proof may extend beyond outcomes'

means—the sole focus of the present paper. What is the relation, for example, between outcomes' *variance* and difficulty of reaching? If nonresponders' outcomes vary more than responders' outcomes, statistics that measure inequality may be affected too. We leave this question to future research.

Once the burden of proof lies with researchers, we trust that the ensuing demand for higher quality paradata, including difficulty-of-reaching measures, will hasten the process of these measures getting better, cleaner, and more widely available.

APPENDIX A

APPENDIX OF CHAPTER 1: ARE ALL INCOME TAXES REALLY THE SAME? AN EXPERIMENTAL INVESTIGATION INTO STATE AND FEDERAL INCOME-TAX SALIENCE

A.1 Simulation

As discussed, reporting error is a potential concern when measuring salience. This concern is potentially magnified when the rate that is reported with more (relative) error is also the rate that is hypothesized to be the less salient rate. This is a concern because reporting error could potentially be driving the finding of lower salience. To explore the validity of this concern, I conducted a series of simulation analyses. To see the extent to which reporting error alone could drive my findings, I simulated the following model:

$$\tau_{F_i} = \tau^*_{F_i} + \epsilon_{F_i} \tag{A.1}$$

$$\tau_{S_i} = \tau^*_{S_i} + \epsilon_{S_i} \tag{A.2}$$

$$\tau_{T_i} = \tau_{S_i} + \tau_{F_i} + \epsilon_{T_i} \tag{A.3}$$

This simulation used the true tax rates from the underlying sample, but then generated reported data using various combinations of model parameters. The simulation focused on matching the error and covariance structure of the data. The errors were drawn from distributions chosen to match the relevant distributions for each reporting rate. In particular, total and federal errors were drawn from a Beta distribution with parameters estimated from the sample population and the state

errors were drawn from a Weibull distribution with similarly estimated parameters. The correlation structure between the three types of error was estimated from the sample data as well. Errors were then generated with the same correlational structure (and alternatively with no correlation and with only federal and state error correlation).

This process was able to create data whose regressions fit the same basic patterns as the actual sample. Looking at Table A.1, I see that the simulation process was able to create data with generally similar relationships between coefficients as in the actual data. Further experimentation with the simulation showed that a potential explanation for this strange pattern of the coefficients for actual and reported state rates is the error of state rates. As can be seen in column 4 of Tables A.2 and A.3, the coefficient for reported state rate decreases and increases when the state error is doubled and halved, respectively; while the other coefficients remain relatively stable.

Table A.1: Simulation regressions, full correlation

VARIABLES	(1) Reported state	(2) Reported fed	(3) P2	(4) P2
State rate, actual	1.065***		0.925***	
	(0.112)		(0.130)	
Federal rate, actual		0.967***	0.947***	
		(0.027)	(0.031)	
State rate, reported				0.117***
				(0.016)
Federal rate, reported				0.767***
				(0.014)
Constant	0.152***	0.099***	0.080***	0.053***
	(0.005)	(0.006)	(0.007)	(0.005)
Observations	3,182	3,182	3,182	3,182
R-squared	0.028	0.290	0.297	0.581

Table A.2: Simulation regressions, doubled state error

VARIABLES	(1) Reported state	(2) Reported fed	(3) P2	(4) P2
State rate, actual	1.131***		0.925***	
	(0.224)		(0.130)	
Federal rate, actual		0.967***	0.947***	
		(0.027)	(0.031)	
State rate, reported				0.049***
				(0.008)
Federal rate, reported				0.775***
				(0.014)
Constant	0.305***	0.099***	0.080***	0.056***
	(0.011)	(0.006)	(0.007)	(0.004)
Observations	3,182	3,182	3,182	3,182
R-squared	0.008	0.290	0.297	0.579

Table A.3: Simulation regressions, halved state error

VARIABLES	(1) Reported state	(2) Reported fed	(3) R_2	(4) R_2
State rate, actual	1.033***		0.925***	
	(0.056)		(0.130)	
Federal rate, actual		0.967***	0.947***	
		(0.027)	(0.031)	
State rate, reported				0.293***
				(0.032)
Federal rate, reported				0.751***
				(0.014)
Constant	0.076***	0.099***	0.080***	0.046***
	(0.003)	(0.006)	(0.007)	(0.005)
Observations	3,182	3,182	3,182	3,182
R-squared	0.097	0.290	0.297	0.586

A.2 Survey text

S1: You are being asked to take part in a research study to help us better understand people's knowledge of the income tax system in the United States. If you agree to participate in this study, we will first ask you a few questions about your household's tax filing behavior to determine if you are eligible. Then we will use this information to generate a fictional household that is similar to yours in many respects. We will ask you some questions about the amount of taxes this fictional household would pay in different situations. You are welcome to use a calculator to

help you answer. Finally, we will ask some more questions about your household, to help us better understand your answers.

We do not anticipate any risks to you for participating in this study other than those encountered in day-to-day life and you may withdraw from the study at any time.

If you are eligible for our survey and complete it, then you will receive a $1 participation fee, and you can also earn up to $0.50 in bonus payments based on how close your answer is to our estimate of the correct answer. At some point in the survey there will be one or more questions to see if you are paying attention and if you do not correctly answer them, you will not be eligible for the bonus payment. There are no other benefits to you for participation in this survey.

Participation in this study is completely voluntary. Your answers will be confidential and all records will be kept private. In any sort of report we make public, we will not include any information that will make it possible to identify you. If you have questions about this survey, the researcher conducting it is Daniel Reeves (dbr88@cornell.edu), a graduate student at Cornell University. If you have any questions or concerns regarding your rights as a subject in this study, you may reach Cornell University's Institutional Review Board at 607-255-5138 or http://www.irb.cornell.edu.

S2: Before we begin, please enter your mturk id, which will be used for your payment.

S3: The tax system has a variety of different tax brackets. The tax rates that households pay depend on which tax bracket they are in. This tax bracket depends on a variety of factors such as the household's income, the deductions they take,

and the number of dependents in the household. To make our questions relevant to your household and its tax bracket, we would first like to ask you some questions about your household's income and taxes.

S4: What state did you live in during 2016? (If you lived in multiple states, pick the one where you spent the majority of the year)

S5: In 2016, what was your total household income rounded to the nearest thousand dollars? *Answers: 0 to 250,000+ (251 levels, in $1,000 increments.)*

S6: How confident are you that your answer for household income is correct? *Answers: Not at all confident...Very confident (7 levels)*

S7: For 2016, how will your household file its income taxes? *Answers: Individual, Head of household or qualifying widow(er), As a married individual (filing jointly), Married individual (filing separately), Don't know.*

S8: For 2016, how many dependents aged 18 or under will your household claim when filing its taxes? (Please enter 99 if you are not sure)

S9: For 2016, how many dependents aged 19 or over will your household claim when filing its taxes? (Please enter 99 if you are not sure)

S10: In 2016, will anyone claim you as a dependent? *Answers: Yes, No, Not sure*

S11: [If married] In 2016, will anyone claim your spouse as a dependent? *Answers: Yes, No, Not sure*

[If answered every question with no unsure/don't know then proceed. Otherwise exit survey]

S12: Thanks. You are eligible to participate in our survey!

S13: For most of this survey, we will ask you about the income taxes paid by the Lawson household. Please carefully read the Lawson household's description as we will test you on this information before you may continue. The Lawson household was chosen to share some features with your household.

The Lawson household:

- Lives in [state]
- Files taxes [filing status]
- Claims X personal exemptions, X for dependents aged 18 or under, and X for dependents over the age of 18. It also has no blind or members aged 65 or over.
- Claims only the standard deductions and exemptions and uses only credits available to it based on its income and dependents.
- Does not live in a city or county with a local income tax
- Receives all its income as salary from one household member's job and has no self-employed members
- Does not do things such as itemizing deductions or investing in tax deductible retirement accounts which might change their taxes
- The Lawson household's income may change from question to question and will be given to you each time.

The Lawson household's income may change from question to question and will be given to you each time.

S14: What state does the Lawson household live in?*Answers: 50 states + DC + Don't know*

S15: Which best describes how the Lawson household earns money? *Answers: The household's income comes from stock investments; The household's income comes from a self-employed member's job; The household's income comes from a government program like Social Security or Disability Insurance; The household's income comes from a single member's job and that member is not self employed; The household's income comes from multiple member's jobs and those members are not self employed; The household's income comes from multiple member's jobs and those members are self employed*

[If answered incorrectly, display message and repeat Lawson section up to one more time; if they fail again then exit survey.]

S16: Unfortunately, you did not correctly answer our questions on the Lawson household. Please read the description and try again.

[If answered correctly, proceed]

S17: Congratulations, you correctly answered our questions about the Lawson household.

This survey contains questions about tax rates and each of these questions has a chance to determine the final amount of your bonus payment. One of these twenty-nine questions will be randomly selected and your bonus payment will depend on how close your answer is to our estimate of the correct answer. Many questions may seem similar, and others are repeated multiple times, so please carefully read each question to make sure you are answering it as accurately as possible. Doing

so will increase the odds that you will receive a higher bonus payment.

Additionally, there will be one, or more, questions to check whether you are paying attention when answering questions. These do not require any tax knowledge, but must be answered correctly in order for you to be eligible for a bonus payment.

S18: We will begin with two training questions to help you learn how to answer the types of questions you will encounter in this survey. The two training questions will ask you to use a randomly chosen tax rate for a fictional household. After the training is complete, you will use your best guess about the actual taxes paid by a household that looks like yours (the Lawson household) to answer the rest of the questions. The training will also explain how your bonus payment is calculated

[Repeat section twice, once with a random rate between 0-50 and another between 50 and 100; random order]

S19: This is a training question, for this training question we have randomly chosen a number between 0 and 100 to be the fictional Felton household's tax rate. The randomly chosen tax rate for the Felton household is [random rate]%. Therefore, the Felton household needs to pay [random high rate]% of its additional earnings in total income taxes.

Imagine that the Felton household's income in 2016 was initially [HH income]. How much more would the Felton household have, after paying all income taxes on the additional earnings, if it had made an additional $100?

S20: Given the randomly chosen tax rate was [random rate]%, the correct answer is [correct answer]. If the Felton household earned $100 more before taxes,

it would need to pay [correct answer]% of that in income taxes, which is [correct answer]. So $100 more before taxes minus [random high rat] in total income taxes is equal to [correct answer] after taxes.

If this had been a question chosen for payment and you gave the correct answer ([correct answer], then you would receive the full bonus payment of $0.50. However, if you answered another number, the further away this number was from [correct answer], the smaller your bonus payment would be if the question were chosen. For instance, if you had answered [wrong answer], then your bonus payment would only be $0.05. If you get too far away from the correct answer, your bonus payment will become $0.00.

Click here to see the formula that will determine your bonus payment. [Appears when clicked] Your bonus payment is the higher of $0 or (1-10*(Correct rate-Income tax rate implied by your answer)2)*$0.50. So your payment is highest ($0.50) at the correct answer and gets smaller as you go above or below the correct answer, until the payment shrinks to zero.

S21: You are now done with the training questions. From here on, you will be answering questions about the Lawson household and will no longer be given a hypothetical income tax rate. Instead, we want you to answer the questions using your best guess about the Lawson household's taxes. There will be six sections of questions, and each section will begin with brief instructions for that section. Any one of the questions about tax rates may be randomly chosen to determine your bonus payment, so please read and answer carefully.

As a reminder, remember that the Lawson household:

- Lives in [state]

- Files taxes [filing status]
- Claims X personal exemptions, X for dependents aged 18 or under, and X for dependents over the age of 18. It also has no blind or members aged 65 or over.
- Claims only the standard deductions and exemptions and uses only credits available to it based on its income and dependents.
- Does not live in a city or county with a local income tax
- Receives all its income as salary from one household member's job and has no self-employed members
- Does not do things such as itemizing deductions or investing in tax deductible retirement accounts which might change their taxes
- The Lawson household's income may change from question to question and will be given to you each time.

S22: For each question in this section we will give you a different starting income for the Lawson household. We then want you to think about all the income taxes the Lawson household would need to pay if it earned an additional $100 and how much of the $100 would be left after it pays all these taxes.

[P1b through Pu15 (5 questions) were presented in random order]

S23: The Lawson household's income in 2016 was initially [HH Income]. How much more would the Lawson household have, after paying **all income taxes** on the additional earnings, if it had made an additional $100?

[From here on out the respondent always had the option to click a button to remind themselves of the basic characteristics of the Lawson household]

S24: The Lawson household's income in 2016 was initially [HH Income*1.05]. How much more would the Lawson household have, after paying **all income taxes** on the additional earnings, if it had made an additional $100?

S25: The Lawson household's income in 2016 was initially [HH Income*1.15]. How much more would the Lawson household have, after paying **all income taxes** on the additional earnings, if it had made an additional $100?

S26: The Lawson household's income in 2016 was initially [HH Income*0.95]. How much more would the Lawson household have, after paying **all income taxes** on the additional earnings, if it had made an additional $100?

S27: The Lawson household's income in 2016 was initially [HH Income*0.85]. How much more would the Lawson household have, after paying **all income taxes** on the additional earnings, if it had made an additional $100?

S28: The Lawson household's income in 2016 is not given. Please ignore the rest of the question and enter Lawson as the answer to show that you are paying attention.

[Randomized to one of three conditions: control, state, federal; with the appropriate words changed depending on randomization]

S29: For each question in this section, we will give you a different starting point of income for the Lawson household. These questions are similar to ones you have been asked before, and we want you to think about **all/state/federal income taxes** the Lawson household would need to pay on its additional income if it earned an additional $100 and how much of the $100 would be left after it pays its all income taxes. However, these questions may be at different income

levels than before.

[P1b through Pu15 (5 questions) were presented in random order]

S30: The Lawson household's income in 2016 was initially [HH Income]. How much more would the Lawson household have, after paying **all/state/federal income taxes** on the additional earnings, if it had made an additional $100?

[From here on out the respondent always had the option to click a button to remind themselves of the basic characteristics of the Lawson household]

S31: The Lawson household's income in 2016 was initially [HH Income*1.03]. How much more would the Lawson household have, after paying **all/state/federal income taxes** on the additional earnings, if it had made an additional $100?

S32: The Lawson household's income in 2016 was initially [HH Income*1.1]. How much more would the Lawson household have, after paying **all/state/federal income taxes** on the additional earnings, if it had made an additional $100?

S33: The Lawson household's income in 2016 was initially [HH Income*0.97]. How much more would the Lawson household have, after paying **all/state/federal income taxes** on the additional earnings, if it had made an additional $100?

S34: The Lawson household's income in 2016 was initially [HH Income*0.9]. How much more would the Lawson household have, after paying **all/state/federal income taxes** on the additional earnings, if it had made an additional $100?

S35: Think back to your answers about how much the Lawson household would pay in **all/state/federal income taxes** at various incomes. Overall, how confident are you that your answers were correct? *Answers: Not at all confident....Very*

confident (7 levels)

S36: Based on your answer, when the Lawson household's income is [HH Income], the Lawson household would pay [answer]% of each additional dollar earned in state income tax. **Using the rate you gave**, if the Lawson household had earned an additional $600 in salary, how much of that $600 would they have left after paying their all/state/federal income tax?

[p1b through pu15 repeated, in random order for all respondents]

S37: When answering the last five questions about all income taxes, what taxes were you considering in your answers? (Check all that apply) *Answers [Random order]: Federal income taxes; State income taxes; City income taxes; Other local income taxes (like county or school district); Payroll taxes (Medicare/Social Security); Other (please specify)*

[Other two sections the respondent was not randomized to were repeated, minus the calculation question]

S38: As our final three questions about the Lawson household, we would like to ask you about the total amounts of each type of income tax that the Lawson household would pay on their income.

[Three average questions presented in random order]

S39: If the Lawson household's income in 2016 was [HH income], what is the total amount they paid in state income taxes?

S40: If the Lawson household's income in 2016 was [HH income], what is the total amount they paid in federal income taxes?

S41: If the Lawson household's income in 2016 was [HH income], what is the total amount they paid in total income taxes?

S42: Finally, we would like to ask you a little bit more about yourself, to help better understand your answers. What year were you born?

S43: What is your gender?

S44: What is the highest level of education you have completed? *Answers: Some high school; High school degree or equivalent; Some college or technical school; Bachelor's degree; Graduate or professional degree; Prefer not to answer*

S45: How will you prepare your federal and state income tax returns for 2016?*Answers: Prepared by you by hand; Prepared by you using tax software; Prepared by another member of your household (either by hand or using tax software); Prepared by a tax professional such as an accountant; Do not know; Prefer not to answer*

S46: Have you ever prepared your own income tax returns, either by hand or using tax software?

S47: In 2016, how many elderly and/or blind exemptions will your household claim? (Please enter 99 if you are not sure)

S48: Does your household claim any tax credits or deductions that are subject to income limits such as the Earned Income Tax Credit (EITC)? *Answers: Yes, EITC only; Yes, some other tax credit or deduction; Yes, EITC and other tax credits; No; Do not know; Prefer not to answer*

S49: For your household's federal income tax returns for 2016, will your house-

hold itemize or take the standard deduction?

S50: What was the approximate total of the itemized deductions that your household took from your federal returns?

S51: Will your household pay the federal alternative minimum tax (AMT) in 2016?

S52: About how many hours do you work per week?

S53: How many members of your household (including you) would you consider to be part-time employees?

S54: Are you or any other members of your household self-employed?

S55 How many members of your household (including yourself) are self-employed?

S56: What is your household's primary source of income? *Answers: Salaries, wages or tips; Self-employment income; Government programs such as Social Security or Disability Insurance; Investment income (e.g. stocks, bonds, property); Other (please specify); Prefer not to answer*

S57: Which would you say best describes your household's living arrangements? *Answers: We own the place that we live in, but are still paying off its mortgage; We own the place that we live in and and have paid off its mortgage: We rent the place we live in; We live rent free at a place we do not own; Other (please describe); Prefer not to answer*

S58: Which term best describes your political affiliation? *Answers: Republican; Democrat; Independent; Libertarian; Green; Other (please describe); Prefer*

not to answer

S59: In a typical week, how much time would you estimate that you spend reading about, listening to, or watching news coverage?

S60: How many years have you lived in [state]?

S61: Where did you live in before living in [state]?

S62: In the past three years, do you think your household has changed which state and/or federal income tax bracket it is in? *Answers: Yes, changed only federal income tax brackets; Yes, changed only state income tax brackets; Yes, changed both state and federal income tax brackets; No, not changed either bracket; Not sure; Prefer not to answer*

S63: When you last moved locations or changed jobs did you use any tools or resources to see how your income taxes might change as a result of this job or location change? *Answers: Yes, I looked at online tax information or used an online calculator; Yes, I consulted an accountant or financial professional; Yes, I talked to a knowledgeable friend or family member; Yes, other (please specify); No, I thought my taxes would change but did not try to determine how much; No, I did not think my taxes would change and therefore did not consult any other resources; No, I did not think about taxes when changing jobs or locations; Not applicable*

S64: What zipcode does your household live in?

APPENDIX B

APPENDIX OF CHAPTER 3: DIFFICULTY TO REACH RESPONDENTS AND NONRESPONSE BIAS: EVIDENCE FROM LARGE GOVERNMENT SURVEYS

Note B.1: Regressor Category Definitions

CPS

Marital status: 6 categories: Married with spouse present, Married with spouse absent, Widowed, Divorced, Separated, Never married.

Number of persons living in the household: 5 categories: 1 person, 2, 3, 4, 5 or more.

State: 51 categories: 50 states and District of Columbia.

Urban status: 2 categories (+ a "Missing" indicator for missing data): Metropolitan, Nonmetropolitan.

Interview month: 12 categories: January, February, . . . , December.

Interview year: 2 categories: 2012, 2013.

BRFSS

Household income: 4 categories: "Is your annual household income from all sources" Less than $25,000, $25,000–49,999, $50,000–$74,999, $75,000 or more.

Marital status: 6 categories (+ Missing): Married, Divorced, Widowed, Separated, Never married, A member of an unmarried couple.

Urban status: 5 categories (+ Missing): In the center city of an Metropolitan Statistical Area (MSA), Outside the center city of an MSA but inside the county containing the center city, Inside a suburban county of the MSA, In an MSA that has no center city, Not in an MSA.

Location: 53 categories: 50 states, District of Columbia, Guam, Puerto Rico.

Interview month: 12 categories: January, February, . . . , December.

CEX

Income: 4 categories ("CEX created variable for the income class of the consumer unit based on income before taxes (created from responses to the detailed income section of the survey)"): Less than $20,000, $20,000–39,999, $40,000–69,999, $70,000 or more.

Marital status: 5 categories ("Marital status of reference person"): Married, Widowed, Divorced, Separated, Never married.

Size of consumer unit: 3 categories: 1 member, 2 members, 3 or more members.

Urban status: 2 categories (+ Missing): Inside metropolitan statistical area (MSA), Outside MSA.

Interview month: 12 categories: January, February, . . . , December.

Interview year: 6 categories: 2008, 2009, 2010, 2011, 2012, 2013.

Table B.1: CPS Demographics, with income

Attempts		1	2	3+	NR	All
Age:	16–19 (%)	6.4	7.0	7.1	6.1	6.5
		(0.1)	(0.1)	(0.1)	(0.2)	(0.0)
	20–39	30.7	33.4	37.0	28.4	31.6
		(0.1)	(0.2)	(0.3)	(0.3)	(0.1)
	40–49	16.5	18.1	18.8	17.3	17.1
		(0.1)	(0.2)	(0.2)	(0.2)	(0.1)
	50–64	26.2	25.7	24.7	28.9	26.2
		(0.1)	(0.2)	(0.2)	(0.3)	(0.1)
	65 and up	20.2	15.8	12.5	19.3	18.6
		(0.1)	(0.2)	(0.2)	(0.3)	(0.1)
Children in household		25.5	28.4	28.5	23.4	26.2
		(0.1)	(0.2)	(0.2)	(0.3)	(0.1)
Female		52.2	52.0	52.0	52.8	52.2
		(0.1)	(0.2)	(0.3)	(0.3)	(0.1)
Educ:	Less than high school	14.9	14.7	14.1	11.1	14.5
		(0.1)	(0.2)	(0.2)	(0.2)	(0.1)
	High school	30.2	29.1	28.0	27.1	29.6
		(0.1)	(0.2)	(0.2)	(0.3)	(0.1)
	Some college or tech. school	27.3	28.0	28.3	27.0	27.5
		(0.1)	(0.2)	(0.2)	(0.3)	(0.1)
	College graduate	27.6	28.2	29.6	34.8	28.5
		(0.1)	(0.2)	(0.3)	(0.3)	(0.1)
Inc:	Less than $20,000	16.8	15.4	15.2	12.5	16.1
		(0.1)	(0.2)	(0.2)	(0.2)	(0.1)
	$20,000-39,999	22.6	22.0	21.4	20.0	22.2
		(0.1)	(0.2)	(0.2)	(0.3)	(0.1)
	$40,000-74,999	27.8	28.0	28.6	28.1	27.9
		(0.1)	(0.2)	(0.2)	(0.3)	(0.1)
	$75,000 and up	32.8	34.7	34.8	39.4	33.9
		(0.1)	(0.2)	(0.3)	(0.3)	(0.1)
Race:	White	82.9	81.1	78.8	84.8	82.3
		(0.1)	(0.2)	(0.2)	(0.2)	(0.1)
	Black	9.7	10.1	11.5	8.6	9.8
		(0.1)	(0.1)	(0.2)	(0.2)	(0.1)
	Asian	4.4	5.6	6.4	4.4	4.8
		(0.0)	(0.1)	(0.1)	(0.1)	(0.0)
	Other	3.1	3.2	3.3	2.3	3.1
		(0.0)	(0.1)	(0.1)	(0.1)	(0.0)
L.F.P.:	Employed	57.9	62.5	67.5	63.5	60.2
		(0.1)	(0.2)	(0.3)	(0.3)	(0.1)
	Unemployed	5.1	5.2	4.8	4.0	5.0
		(0.0)	(0.1)	(0.1)	(0.1)	(0.0)
	Not in the labor force	37.0	32.3	27.7	32.5	34.9
		(0.1)	(0.2)	(0.2)	(0.3)	(0.1)
Labor force participation		63.0	67.7	72.3	67.5	65.1
		(0.1)	(0.2)	(0.2)	(0.3)	(0.1)
Unemployment rate		8.1	7.6	6.7	5.9	7.6
		(0.1)	(0.1)	(0.2)	(0.2)	(0.1)
Median number of attempts (known)		1	2	3	n/a	1
Observations		197,751	52,275	32,770	24,807	307,603

Notes: This table replicates Table 3.1, but includes household income categories.

Table B.2: Labor force participation, self-reported only

Attempts		1	2	3+	NR
		A: Regression with interactions			
		Base	Interactions		
Age:	16–19	-0.176***	-0.030	-0.067*	0.090
		(0.015)	(0.032)	(0.036)	(0.062)
	20–39	-0.001	-0.013*	-0.003	-0.020**
		(0.004)	(0.008)	(0.008)	(0.010)
	50–64	-0.129***	0.019**	0.051***	0.023**
		(0.004)	(0.009)	(0.009)	(0.011)
	65 and up	-0.568***	0.014	0.062***	0.022
		(0.005)	(0.011)	(0.014)	(0.014)
Children in household		0.016***	0.008	0.014	0.004
		(0.004)	(0.009)	(0.010)	(0.012)
Female		-0.104***	-0.001	0.021***	0.019***
		(0.003)	(0.006)	(0.006)	(0.007)
Educ:	Less than high school	-0.113***	0.009	-0.007	-0.030*
		(0.005)	(0.011)	(0.013)	(0.016)
	Some college or tech. school	0.044***	-0.010	-0.017**	-0.016
		(0.003)	(0.008)	(0.009)	(0.010)
	College graduate	0.114***	-0.018**	-0.039***	-0.017*
		(0.003)	(0.007)	(0.008)	(0.010)
Race:	Black	-0.015***	-0.009	0.008	-0.022
		(0.005)	(0.010)	(0.011)	(0.014)
	Asian	-0.038***	0.016	-0.001	0.015
		(0.007)	(0.014)	(0.016)	(0.019)
	Other	-0.031***	-0.013	0.025	0.012
		(0.008)	(0.018)	(0.020)	(0.027)
Constant		0.725***	-0.002	0.140***	0.097*
		(0.014)	(0.032)	(0.034)	(0.051)
		B: Adjusted means			
Labor force participation		0.624***	0.659***	0.699***	0.677***
		(0.001)	(0.003)	(0.003)	(0.004)

Notes: This table replicates Table 3.2, but restricts the sample to self reports. $N = 155,442$ (1 attempt: 99,604; 2: 25,813; 3+: 16,675; None Reported: 13,350). $R^2 = 0.32$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.3: Labor force participation, proxy-reported only

Attempts	1	2	3+	NR
	A: Regression with interactions			
	Base	Interactions		
Age: 16–19	-0.255***	-0.013	-0.025	-0.037*
	(0.007)	(0.015)	(0.018)	(0.022)
20–39	-0.015***	-0.001	-0.004	-0.022*
	(0.004)	(0.008)	(0.010)	(0.012)
50–64	-0.096***	0.016*	0.043***	0.038***
	(0.004)	(0.009)	(0.011)	(0.013)
65 and up	-0.580***	0.020	0.077***	0.049***
	(0.005)	(0.012)	(0.016)	(0.017)
Children in household	0.046***	0.015*	0.031***	0.023*
	(0.004)	(0.009)	(0.011)	(0.014)
Female	-0.103***	-0.005	-0.004	-0.009
	(0.003)	(0.006)	(0.007)	(0.008)
Educ: Less than high school	-0.155***	0.007	0.007	-0.013
	(0.005)	(0.010)	(0.013)	(0.016)
Some college or tech. school	0.018***	-0.002	-0.009	0.002
	(0.004)	(0.008)	(0.010)	(0.012)
College graduate	0.086***	-0.012	-0.014	0.009
	(0.004)	(0.008)	(0.009)	(0.011)
Race: Black	-0.042***	0.005	-0.006	-0.016
	(0.005)	(0.012)	(0.013)	(0.017)
Asian	-0.072***	-0.008	0.027*	0.009
	(0.007)	(0.014)	(0.016)	(0.022)
Other	-0.042***	0.004	0.004	0.080***
	(0.008)	(0.018)	(0.023)	(0.027)
Constant	0.756***	-0.011	0.132*	0.102
	(0.032)	(0.067)	(0.072)	(0.076)
	B: Adjusted means			
Labor force participation	0.658***	0.665***	0.681***	0.656***
	(0.001)	(0.003)	(0.004)	(0.004)

Notes: This table replicates Table 3.2, but restricts the sample to proxy reports. $N = 149,685$ (1 attempt: 96,891; 2: 25,982; 3+: 15,674; None Reported: 11,138). $R^2 = 0.27$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.4: Unemployment rate, self-reported only

Attempts	1	2	3+	NR
	A: Regression with interactions			
	Base	Interactions		
Age: 16–19	0.116***	0.003	0.051	-0.016
	(0.017)	(0.037)	(0.043)	(0.065)
20–39	0.018***	-0.003	-0.011*	-0.015**
	(0.003)	(0.006)	(0.006)	(0.007)
50–64	0.009***	-0.003	-0.008	-0.003
	(0.003)	(0.006)	(0.007)	(0.007)
65 and up	0.012**	0.001	-0.015	0.013
	(0.005)	(0.010)	(0.011)	(0.012)
Children in household	-0.012***	0.003	0.005	0.009
	(0.004)	(0.008)	(0.008)	(0.009)
Female	-0.001	-0.002	0.002	0.002
	(0.002)	(0.005)	(0.005)	(0.005)
Educ: Less than high school	0.047***	0.007	-0.020	-0.009
	(0.006)	(0.012)	(0.013)	(0.018)
Some college or tech. school	-0.015***	0.009	-0.002	0.008
	(0.003)	(0.006)	(0.007)	(0.008)
College graduate	-0.047***	0.008	0.008	0.015**
	(0.003)	(0.006)	(0.006)	(0.007)
Race: Black	0.058***	-0.015	-0.014	-0.005
	(0.005)	(0.009)	(0.010)	(0.013)
Asian	-0.010*	0.011	0.013	-0.016
	(0.005)	(0.011)	(0.011)	(0.011)
Other	0.044***	-0.005	0.016	0.000
	(0.008)	(0.016)	(0.019)	(0.024)
Constant	0.063***	-0.028	-0.007	-0.020
	(0.014)	(0.026)	(0.026)	(0.040)
	B: Adjusted means			
Unemployment rate	0.084***	0.072***	0.059***	0.060***
	(0.001)	(0.002)	(0.002)	(0.003)

Notes: This table replicates Table 3.3, but restricts the sample to self reports. $N = 99,752$ (1 attempt: 60,524; 2: 17,585; 3+: 12,499; None Reported: 9,144). $R^2 = 0.04$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.5: Unemployment rate, proxy-reported only

Attempts	1	2	3+	NR
	A: Regression with interactions			
	Base	Interactions		
Age: 16–19	0.076***	0.021	0.010	-0.025
	(0.007)	(0.016)	(0.019)	(0.023)
20–39	0.011***	0.008	-0.012*	0.003
	(0.003)	(0.006)	(0.007)	(0.009)
50–64	-0.006**	-0.004	-0.013*	0.004
	(0.003)	(0.006)	(0.007)	(0.008)
65 and up	-0.011***	0.001	0.004	0.010
	(0.004)	(0.010)	(0.013)	(0.012)
Children in household	-0.008**	-0.009	-0.004	0.002
	(0.003)	(0.007)	(0.008)	(0.010)
Female	-0.009***	-0.007	-0.002	0.012**
	(0.002)	(0.005)	(0.005)	(0.006)
Educ: Less than high school	0.041***	-0.010	-0.006	0.008
	(0.005)	(0.011)	(0.012)	(0.018)
Some college or tech. school	-0.029***	-0.002	0.004	0.000
	(0.003)	(0.006)	(0.007)	(0.009)
College graduate	-0.041***	-0.001	0.008	-0.000
	(0.003)	(0.005)	(0.006)	(0.007)
Race: Black	0.068***	-0.020*	-0.004	-0.012
	(0.005)	(0.011)	(0.012)	(0.016)
Asian	-0.004	0.003	0.005	-0.030**
	(0.005)	(0.009)	(0.012)	(0.013)
Other	0.051***	-0.020	-0.008	0.003
	(0.008)	(0.017)	(0.021)	(0.026)
Constant	0.070*	-0.015	-0.109**	-0.153***
	(0.039)	(0.077)	(0.045)	(0.056)
	B: Adjusted means			
Unemployment rate	0.077***	0.078***	0.070***	0.073***
	(0.001)	(0.002)	(0.003)	(0.003)

Notes: This table replicates Table 3.3, but restricts the sample to proxy reports. $N = 98,887$ (1 attempt: 63,144; 2: 17,454; 3+: 10,909; None Reported: 7,380). $R^2 = 0.06$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.6: Labor force participation, telephone interviews only

Attempts		1	2	3+	NR
		A: Regression with interactions			
		Base	Interactions		
Age:	16–19	-0.243***	0.013	-0.046	-0.068**
		(0.020)	(0.040)	(0.046)	(0.027)
	20–39	-0.018**	-0.011	0.002	-0.016
		(0.008)	(0.016)	(0.018)	(0.011)
	50–64	-0.084***	0.002	0.036*	-0.004
		(0.009)	(0.017)	(0.020)	(0.012)
	65 and up	-0.554***	0.047*	0.099***	0.006
		(0.013)	(0.025)	(0.031)	(0.016)
Children in household		0.070***	0.001	-0.010	-0.007
		(0.010)	(0.019)	(0.021)	(0.013)
Female		-0.090***	-0.013	-0.002	-0.006
		(0.006)	(0.011)	(0.013)	(0.008)
Educ:	Less than high school	-0.148***	-0.018	-0.002	-0.011
		(0.013)	(0.027)	(0.031)	(0.017)
	Some college or tech. school	0.014	0.012	0.023	0.010
		(0.009)	(0.016)	(0.019)	(0.011)
	College graduate	0.080***	0.002	0.021	0.016
		(0.008)	(0.016)	(0.018)	(0.011)
Race:	Black	-0.038***	0.007	0.017	-0.005
		(0.013)	(0.026)	(0.026)	(0.017)
	Asian	-0.024	-0.041	0.011	-0.023
		(0.016)	(0.029)	(0.032)	(0.021)
	Other	-0.046**	0.045	0.021	0.065**
		(0.020)	(0.041)	(0.048)	(0.028)
Constant		0.756***	0.206***	0.097	0.142***
		(0.038)	(0.070)	(0.079)	(0.053)
			B: Adjusted means		
Labor force participation		0.685***	0.703***	0.713***	0.684***
		(0.003)	(0.006)	(0.007)	(0.003)

Notes: This table replicates Table 3.2, but restricts the sample to telephone interviews $N = 52,624$ (1 attempt: 17,917; 2: 6,277; 3+: 4,315; None Reported: 24,115). $R^2 = 0.28$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.7: Unemployment rate, telephone interviews only

Attempts	1	2	3+	NR
	A: Regression with interactions			
	Base	Interactions		
Age: 16–19	0.049**	0.044	0.063	0.018
	(0.019)	(0.040)	(0.048)	(0.027)
20–39	0.023***	-0.010	-0.031**	-0.014*
	(0.006)	(0.011)	(0.012)	(0.008)
50–64	0.010*	-0.010	-0.010	-0.006
	(0.006)	(0.011)	(0.013)	(0.008)
65 and up	0.004	-0.014	-0.013	0.014
	(0.009)	(0.018)	(0.020)	(0.012)
Children in household	-0.005	-0.012	-0.003	-0.005
	(0.007)	(0.014)	(0.015)	(0.010)
Female	-0.004	-0.002	0.006	0.004
	(0.004)	(0.008)	(0.009)	(0.006)
Educ: Less than high school	0.057***	-0.013	-0.001	-0.010
	(0.014)	(0.026)	(0.029)	(0.019)
Some college or tech. school	-0.022***	0.011	0.021	0.004
	(0.006)	(0.012)	(0.013)	(0.008)
College graduate	-0.042***	0.016	0.023**	0.006
	(0.006)	(0.010)	(0.012)	(0.007)
Race: Black	0.059***	-0.029	-0.048**	-0.001
	(0.012)	(0.020)	(0.020)	(0.016)
Asian	-0.024***	0.000	0.036	-0.008
	(0.009)	(0.018)	(0.025)	(0.012)
Other	0.059***	-0.042	-0.044	-0.009
	(0.019)	(0.034)	(0.034)	(0.025)
Constant	0.046*	-0.039	-0.016	-0.026
	(0.027)	(0.041)	(0.047)	(0.037)
	B: Adjusted means			
Unemployment rate	0.064***	0.060***	0.051***	0.060***
	(0.002)	(0.004)	(0.005)	(0.002)

Notes: This table replicates Table 3.3, but restricts the sample to telephone interviews. $N = 36,281$ (1 attempt: 12,300; 2: 4,523; 3+: 3,185; None Reported: 16,273). $R^2 = 0.05$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.8: Obesity, with controls for labor force participation

Attempts		1	2–3	4–6	7+
			A: Regression with interactions		
		Base	Interactions		
Age:	18–39	-0.073***	0.002	0.001	0.023***
		(0.005)	(0.007)	(0.007)	(0.007)
	40–49	-0.014***	0.012*	-0.002	0.009
		(0.005)	(0.007)	(0.007)	(0.007)
	60–69	-0.001	0.002	0.007	-0.000
		(0.004)	(0.006)	(0.007)	(0.006)
	70 and up	-0.104***	0.008	0.007	0.017**
		(0.005)	(0.007)	(0.008)	(0.008)
Children in household		0.026***	-0.007	-0.004	-0.019***
		(0.004)	(0.005)	(0.006)	(0.006)
Female		-0.002	-0.008**	-0.023***	-0.021***
		(0.003)	(0.004)	(0.004)	(0.004)
Educ:	Less than high school	0.014***	0.002	0.003	0.009
		(0.005)	(0.007)	(0.008)	(0.008)
	Some college or tech. school	-0.004	0.005	-0.003	0.004
		(0.003)	(0.005)	(0.005)	(0.005)
	College graduate	-0.062***	0.003	-0.006	-0.008
		(0.004)	(0.005)	(0.005)	(0.005)
Inc:	Below $25,000	0.019***	-0.002	-0.002	-0.012*
		(0.005)	(0.007)	(0.007)	(0.007)
	$25,000-49,999	0.006	-0.004	-0.004	-0.003
		(0.005)	(0.006)	(0.007)	(0.007)
	$75,000 and up	-0.042***	0.002	-0.001	0.007
		(0.005)	(0.006)	(0.007)	(0.007)
Race:	Black, non-hispanic	0.117***	0.001	-0.009	0.013*
		(0.005)	(0.007)	(0.008)	(0.007)
	Other, non-hispanic	-0.008	0.011	0.023***	0.000
		(0.006)	(0.008)	(0.009)	(0.009)
	Hispanic	0.034***	0.002	0.003	0.003
		(0.007)	(0.009)	(0.009)	(0.009)
Empl:	Self-employed	-0.048***	-0.005	-0.014*	-0.007
		(0.005)	(0.007)	(0.008)	(0.007)
	Out of work, >1 year	0.026***	-0.017	-0.009	0.001
		(0.008)	(0.011)	(0.012)	(0.012)
	Out of work, <1 year	0.001	0.002	-0.024*	-0.017
		(0.008)	(0.011)	(0.012)	(0.013)
	Homemaker	-0.035***	0.001	-0.009	0.005
		(0.006)	(0.008)	(0.009)	(0.009)
	Student	-0.094***	0.016	0.003	0.007
		(0.009)	(0.012)	(0.013)	(0.013)
	Retired	-0.002	0.001	-0.015**	0.001
		(0.004)	(0.006)	(0.007)	(0.007)
	Unable to work	0.106***	0.001	-0.023***	-0.012
		(0.005)	(0.008)	(0.008)	(0.009)
Constant		0.357***	0.015	0.048***	0.002
		(0.012)	(0.016)	(0.017)	(0.018)
			B: Adjusted means		
Obesity		0.296***	0.286***	0.282***	0.268***
		(0.001)	(0.001)	(0.001)	(0.002)

Notes: This table replicates Table 3.5, but includes additional controls for the employment-status question "Are you currently:" (see response options under "Empl" in table; omitted category: "Employed for wages"). $N = 450,212$ (1 attempt: 114,694; 2–3: 137,418; 4–6: 98,813; 7+: 99,287). $R^2 = 0.04$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.9: Self-reported weight

Attempts		1	2–3	4–6	7+
		A: Regression with interactions			
		Base	Interactions		
Age:	18–39	-4.068***	0.240	0.589**	1.550***
		(0.205)	(0.272)	(0.288)	(0.285)
	40–49	-0.293	0.300	-0.144	0.606**
		(0.205)	(0.272)	(0.287)	(0.280)
	60–69	-1.097***	-0.058	0.345	0.374
		(0.170)	(0.228)	(0.248)	(0.243)
	70 and up	-7.152***	0.357	0.597**	1.582***
		(0.177)	(0.241)	(0.267)	(0.268)
Children in household		1.060***	-0.147	-0.069	-0.617***
		(0.160)	(0.212)	(0.224)	(0.223)
Female		-15.797***	-0.340**	-0.721***	-0.916***
		(0.112)	(0.151)	(0.163)	(0.162)
Educ:	Less than high school	-0.102	-0.397	-0.307	-0.317
		(0.203)	(0.278)	(0.305)	(0.306)
	Some college or tech. school	0.616***	0.019	-0.109	0.016
		(0.140)	(0.190)	(0.208)	(0.210)
	College graduate	-1.453***	-0.103	-0.229	-0.352*
		(0.144)	(0.195)	(0.212)	(0.211)
Inc:	Below $25,000	0.760***	-0.143	-0.411	-0.750***
		(0.191)	(0.258)	(0.280)	(0.282)
	$25,000-49,999	-0.071	-0.024	-0.084	-0.112
		(0.185)	(0.248)	(0.269)	(0.270)
	$75,000 and up	-1.506***	0.097	0.205	0.315
		(0.191)	(0.253)	(0.271)	(0.267)
Race:	Black, non-hispanic	5.749***	-0.011	-0.239	0.379
		(0.217)	(0.292)	(0.311)	(0.300)
	Other, non-hispanic	-2.593***	0.459	0.658*	-0.345
		(0.244)	(0.326)	(0.348)	(0.347)
	Hispanic	-1.640***	-0.609*	-0.592	-0.676*
		(0.281)	(0.364)	(0.374)	(0.370)
Constant		94.401***	-0.156	0.202	-0.373
		(0.465)	(0.629)	(0.691)	(0.732)
			B: Adjusted means		
Self-reported weight (kg)		80.385***	79.932***	79.644***	79.067***
		(0.055)	(0.048)	(0.057)	(0.064)

Notes: This table replicates Table 3.5, but replaces the dependent variable with weight. $N = 450,212$ (1 attempt: 114,694; 2–3: 137,418; 4–6: 98,813; 7+: 99,287). $R^2 = 0.2$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.10: Self-reported height

Attempts		1	2–3	4–6	7+
		A: Regression with interactions			
		Base	Interactions		
Age:	18–39	0.010***	-0.001	-0.001	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)
	40–49	0.005***	-0.001	-0.001	0.001
		(0.001)	(0.001)	(0.001)	(0.001)
	60–69	-0.008***	-0.001	0.001	0.002*
		(0.001)	(0.001)	(0.001)	(0.001)
	70 and up	-0.019***	0.000	0.001	0.002*
		(0.001)	(0.001)	(0.001)	(0.001)
Children in household		-0.000	0.001	0.002*	0.001
		(0.001)	(0.001)	(0.001)	(0.001)
Female		-0.149***	0.001	0.003***	0.002***
		(0.000)	(0.001)	(0.001)	(0.001)
Educ:	Less than high school	-0.008***	-0.003***	-0.004***	-0.006***
		(0.001)	(0.001)	(0.001)	(0.001)
	Some college or tech. school	0.007***	-0.000	0.000	0.001
		(0.001)	(0.001)	(0.001)	(0.001)
	College graduate	0.011***	-0.000	0.000	0.001
		(0.001)	(0.001)	(0.001)	(0.001)
Inc:	Below $25,000	-0.008***	-0.001	-0.000	0.000
		(0.001)	(0.001)	(0.001)	(0.001)
	$25,000-49,999	-0.004***	-0.001	0.000	-0.000
		(0.001)	(0.001)	(0.001)	(0.001)
	$75,000 and up	0.004***	-0.000	0.002**	0.002*
		(0.001)	(0.001)	(0.001)	(0.001)
Race:	Black, non-hispanic	0.001	0.001	-0.000	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)
	Other, non-hispanic	-0.023***	-0.002	-0.004**	-0.006***
		(0.001)	(0.001)	(0.001)	(0.001)
	Hispanic	-0.037***	-0.006***	-0.009***	-0.008***
		(0.001)	(0.001)	(0.001)	(0.001)
Constant		1.794***	-0.005**	-0.010***	-0.007**
		(0.002)	(0.003)	(0.003)	(0.003)
			B: Adjusted means		
Height (m)		1.693***	1.693***	1.693***	1.693***
		(0.000)	(0.000)	(0.000)	(0.000)

Notes: This table replicates Table 3.5, but replaces the dependent variable with height. $N = 450,212$ (1 attempt: 114,694; 2–3: 137,418; 4–6: 98,813; 7+: 99,287). $R^2 = 0.54$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.11: Obesity, weighted

Attempts		1	2–3	4–6	7+
		A: Regression with interactions			
		Base	Interactions		
Age:	18–39	-0.106***	0.021	0.012	0.031**
		(0.010)	(0.013)	(0.014)	(0.014)
	40–49	-0.016	0.024*	-0.004	-0.013
		(0.010)	(0.014)	(0.014)	(0.014)
	60–69	-0.007	0.000	0.011	-0.003
		(0.009)	(0.012)	(0.013)	(0.013)
	70 and up	-0.120***	0.014	0.011	0.012
		(0.009)	(0.012)	(0.014)	(0.014)
Children in household		0.022***	-0.003	-0.001	-0.009
		(0.008)	(0.010)	(0.010)	(0.011)
Female		-0.005	-0.007	-0.014*	-0.014*
		(0.005)	(0.007)	(0.008)	(0.008)
Educ:	Less than high school	0.009	0.012	0.007	0.010
		(0.010)	(0.014)	(0.015)	(0.015)
	Some college or tech. school	-0.002	-0.004	-0.009	0.007
		(0.007)	(0.009)	(0.010)	(0.010)
	College graduate	-0.060***	-0.005	-0.004	-0.001
		(0.007)	(0.009)	(0.010)	(0.010)
Inc:	Below $25,000	0.030***	-0.002	-0.011	-0.020
		(0.010)	(0.013)	(0.014)	(0.015)
	$25,000-49,999	-0.005	-0.001	0.003	0.000
		(0.009)	(0.012)	(0.013)	(0.014)
	$75,000 and up	-0.053***	0.011	0.021	0.007
		(0.009)	(0.012)	(0.013)	(0.013)
Race:	Black, non-hispanic	0.125***	-0.017	-0.034**	-0.010
		(0.011)	(0.015)	(0.016)	(0.016)
	Other, non-hispanic	-0.055***	0.006	0.017	-0.018
		(0.011)	(0.015)	(0.015)	(0.015)
	Hispanic	0.048***	-0.022	0.003	-0.016
		(0.013)	(0.017)	(0.017)	(0.017)
Constant		0.408***	-0.020	-0.005	-0.010
		(0.021)	(0.028)	(0.030)	(0.032)
		B: Adjusted means			
Obesity		0.294***	0.284***	0.272***	0.261***
		(0.003)	(0.003)	(0.003)	(0.003)

Notes: This table replicates Table 3.5, but uses the BRFSS analysis weights. $N = 450,212$ (1 attempt: 114,694; 2–3: 137,418; 4–6: 98,813; 7+: 99,287). $R^2 = 0.04$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.12: Self-reported weight, weighted

Attempts		1	2–3	4–6	7+
		A: Regression with interactions			
		Base	Interactions		
Age:	18–39	-4.686***	0.899	0.793	1.757***
		(0.425)	(0.558)	(0.590)	(0.583)
	40–49	0.333	-0.020	-1.089*	-0.976*
		(0.422)	(0.567)	(0.587)	(0.579)
	60–69	-1.022***	0.040	0.069	0.195
		(0.347)	(0.468)	(0.502)	(0.490)
	70 and up	-6.666***	-0.040	0.076	0.879
		(0.366)	(0.488)	(0.563)	(0.538)
Children in household		0.886***	0.037	0.278	-0.068
		(0.333)	(0.444)	(0.468)	(0.465)
Female		-15.455***	-0.426	-0.432	-0.515
		(0.244)	(0.326)	(0.342)	(0.343)
Educ:	Less than high school	-0.720	0.509	-0.377	-0.446
		(0.446)	(0.644)	(0.674)	(0.644)
	Some college or tech. school	1.153***	-0.708*	-0.232	-0.149
		(0.304)	(0.413)	(0.430)	(0.444)
	College graduate	-1.191***	-0.572	-0.143	-0.340
		(0.289)	(0.389)	(0.426)	(0.413)
Inc:	Below $25,000	0.504	-0.243	-0.562	-1.131*
		(0.403)	(0.554)	(0.587)	(0.592)
	$25,000-49,999	-0.131	-0.457	-0.245	-0.170
		(0.374)	(0.503)	(0.534)	(0.562)
	$75,000 and up	-1.230***	-0.178	-0.086	-0.273
		(0.410)	(0.527)	(0.546)	(0.551)
Race:	Black, non-hispanic	5.703***	-0.673	-1.203	-0.402
		(0.565)	(0.719)	(0.741)	(0.736)
	Other, non-hispanic	-5.324***	-1.078	0.214	-2.753***
		(0.475)	(0.661)	(0.677)	(0.662)
	Hispanic	-1.701***	-1.610**	-0.197	-0.908
		(0.566)	(0.742)	(0.741)	(0.708)
Constant		94.169***	-0.294	0.375	1.312
		(0.919)	(1.209)	(1.258)	(1.449)
			B: Adjusted means		
Self-reported weight (kg)		81.127***	80.604***	80.096***	79.623***
		(0.135)	(0.115)	(0.116)	(0.114)

Notes: This table replicates Table B.9, but uses the BRFSS analysis weights. $N = 450,212$ (1 attempt: 114,694; 2–3: 137,418; 4–6: 98,813; 7+: 99,287). $R^2 = 0.2$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.13: Self-reported height, weighted

Attempts		1	2–3	4–6	7+
		A: Regression with interactions			
		Base	Interactions		
Age:	18–39	0.011***	-0.001	-0.003	-0.000
		(0.002)	(0.002)	(0.002)	(0.002)
	40–49	0.006***	-0.003	-0.004*	-0.001
		(0.002)	(0.002)	(0.002)	(0.002)
	60–69	-0.008***	0.002	0.001	0.003
		(0.001)	(0.002)	(0.002)	(0.002)
	70 and up	-0.017***	-0.000	0.001	0.001
		(0.001)	(0.002)	(0.002)	(0.002)
Children in household		-0.003**	0.002	0.005***	0.002
		(0.001)	(0.002)	(0.002)	(0.002)
Female		-0.148***	0.001	0.004***	0.004***
		(0.001)	(0.001)	(0.001)	(0.001)
Educ:	Less than high school	-0.007***	-0.006**	-0.007**	-0.012***
		(0.002)	(0.003)	(0.003)	(0.003)
	Some college or tech. school	0.008***	-0.002	0.003*	0.000
		(0.001)	(0.002)	(0.002)	(0.002)
	College graduate	0.011***	-0.002	0.000	-0.002
		(0.001)	(0.002)	(0.002)	(0.002)
Inc:	Below $25,000	-0.009***	-0.000	-0.002	-0.002
		(0.002)	(0.002)	(0.002)	(0.003)
	$25,000-49,999	-0.003**	-0.002	-0.002	-0.002
		(0.002)	(0.002)	(0.002)	(0.002)
	$75,000 and up	0.006***	-0.000	-0.001	-0.000
		(0.002)	(0.002)	(0.002)	(0.002)
Race:	Black, non-hispanic	0.001	-0.002	-0.002	-0.001
		(0.002)	(0.003)	(0.003)	(0.003)
	Other, non-hispanic	-0.031***	-0.010***	-0.006*	-0.016***
		(0.003)	(0.003)	(0.003)	(0.003)
	Hispanic	-0.036***	-0.012***	-0.015***	-0.012***
		(0.002)	(0.003)	(0.003)	(0.003)
Constant		1.787***	-0.001	-0.001	0.005
		(0.003)	(0.005)	(0.005)	(0.005)
			B: Adjusted means		
Height (m)		1.704***	1.703***	1.703***	1.702***
		(0.001)	(0.000)	(0.000)	(0.001)

Notes: This table replicates Table B.10, but uses the BRFSS analysis weights. $N = 450,212$ (1 attempt: 114,694; 2–3: 137,418; 4–6: 98,813; 7+: 99,287). $R^2 = 0.52$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.14: Obesity, interviewer-ID-controls sample

Attempts		1	2–3	4–6	7+
		A: Regression with interactions			
		Base	Interactions		
Age:	18–39	-0.090***	0.005	0.014	0.033***
		(0.006)	(0.008)	(0.009)	(0.009)
	40–49	-0.016**	0.014*	0.002	0.010
		(0.006)	(0.008)	(0.009)	(0.009)
	60–69	-0.006	0.007	0.007	0.008
		(0.005)	(0.007)	(0.007)	(0.007)
	70 and up	-0.120***	0.012	0.005	0.028***
		(0.005)	(0.007)	(0.008)	(0.008)
Children in household		0.021***	-0.005	0.000	-0.016**
		(0.005)	(0.006)	(0.007)	(0.007)
Female		-0.003	-0.004	-0.025***	-0.014***
		(0.003)	(0.005)	(0.005)	(0.005)
Educ:	Less than high school	0.025***	0.001	-0.003	0.006
		(0.006)	(0.008)	(0.009)	(0.009)
	Some college or tech. school	-0.007*	0.010*	-0.005	0.003
		(0.004)	(0.006)	(0.006)	(0.006)
	College graduate	-0.066***	0.006	-0.004	-0.003
		(0.004)	(0.006)	(0.006)	(0.006)
Inc:	Below $25,000	0.034***	-0.006	-0.006	-0.016*
		(0.006)	(0.008)	(0.009)	(0.009)
	$25,000-49,999	0.008	-0.008	0.000	-0.007
		(0.006)	(0.008)	(0.008)	(0.008)
	$75,000 and up	-0.046***	-0.001	0.003	0.004
		(0.006)	(0.008)	(0.008)	(0.008)
Race:	Black, non-hispanic	0.127***	0.003	-0.007	0.004
		(0.007)	(0.009)	(0.010)	(0.009)
	Other, non-hispanic	-0.005	0.016	0.030***	-0.002
		(0.007)	(0.010)	(0.010)	(0.010)
	Hispanic	0.037***	-0.007	-0.008	-0.003
		(0.008)	(0.011)	(0.011)	(0.011)
Constant		0.371***	-0.001	0.029	-0.004
		(0.013)	(0.018)	(0.020)	(0.021)
		B: Adjusted means			
Obesity		0.298***	0.289***	0.285***	0.266***
		(0.002)	(0.001)	(0.002)	(0.002)

Notes: This table replicates Table 3.5, but limits the sample to only include respondents interviewed by interviewers with at least 10 interviews in each of the four difficulty categories (1,651 interviewers). $N = 303,034$ (1 attempt: 78,998; 2–3: 91,931; 4–6: 66,194; 7+: 65,911). $R^2 = 0.04$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.15: Obesity, interviewer-ID-controls sample with interviewer-ID controls

Attempts		1	2–3	4–6	7+
		A: Regression with interactions			
		Base	Interactions		
Age:	18–39	-0.091***	0.006	0.014	0.035***
		(0.006)	(0.008)	(0.009)	(0.009)
	40–49	-0.016***	0.016*	0.003	0.012
		(0.006)	(0.008)	(0.009)	(0.009)
	60–69	-0.005	0.007	0.007	0.008
		(0.005)	(0.007)	(0.008)	(0.007)
	70 and up	-0.120***	0.010	0.004	0.027***
		(0.005)	(0.007)	(0.008)	(0.008)
Children in household		0.023***	-0.006	-0.001	-0.018***
		(0.005)	(0.007)	(0.007)	(0.007)
Female		-0.003	-0.006	-0.025***	-0.014***
		(0.003)	(0.005)	(0.005)	(0.005)
Educ:	Less than high school	0.026***	-0.000	-0.003	0.006
		(0.006)	(0.008)	(0.009)	(0.009)
	Some college or tech. school	-0.007	0.010*	-0.007	0.002
		(0.004)	(0.006)	(0.006)	(0.006)
	College graduate	-0.065***	0.006	-0.005	-0.003
		(0.004)	(0.006)	(0.007)	(0.007)
Inc:	Below $25,000	0.035***	-0.007	-0.005	-0.019**
		(0.006)	(0.008)	(0.009)	(0.009)
	$25,000-49,999	0.008	-0.009	0.002	-0.007
		(0.006)	(0.008)	(0.008)	(0.008)
	$75,000 and up	-0.046***	-0.001	0.003	0.004
		(0.006)	(0.008)	(0.008)	(0.008)
Race:	Black, non-hispanic	0.128***	0.002	-0.007	0.002
		(0.007)	(0.009)	(0.010)	(0.009)
	Other, non-hispanic	-0.006	0.017*	0.031***	-0.002
		(0.007)	(0.010)	(0.010)	(0.010)
	Hispanic	0.034***	-0.006	-0.007	-0.003
		(0.008)	(0.011)	(0.012)	(0.011)
Constant		0.391***	-0.092**	0.028	-0.017
		(0.031)	(0.045)	(0.050)	(0.055)
		B: Adjusted means			
Obesity		0.298***	0.289***	0.285***	0.266***
		(0.002)	(0.002)	(0.002)	(0.002)

Notes: This table replicates Table 3.5, but limits the sample to only include respondents interviewed by interviewers with at least 10 interviews in each of the four difficulty categories (1,651 interviewers), and includes interviewer-ID controls (and their difficulty interactions). $N = 303,034$ (1 attempt: 78,998; 2–3: 91,931; 4–6: 66,194; 7+: 65,911). $R^2 = 0.06$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.16: Total expenditures, second-interview difficulty

Attempts		1	2	3–4	5+
		A: Regression with interactions			
		Base	Interactions		
Age:	16–29	-0.145***	0.045	0.073**	0.062*
		(0.028)	(0.035)	(0.033)	(0.032)
	30–39	-0.025	-0.011	0.019	-0.032
		(0.025)	(0.031)	(0.029)	(0.029)
	50–64	-0.026	0.008	0.022	0.014
		(0.023)	(0.028)	(0.027)	(0.026)
	65 and up	-0.079***	0.034	0.039	0.021
		(0.025)	(0.032)	(0.030)	(0.031)
Children in household		0.034	0.009	-0.025	0.025
		(0.026)	(0.033)	(0.031)	(0.031)
Female		-0.016	0.017	0.009	0.002
		(0.014)	(0.018)	(0.017)	(0.017)
Educ:	Less than high school	-0.116***	-0.051*	-0.017	0.002
		(0.021)	(0.027)	(0.027)	(0.027)
	Some college or tech. school	0.108***	-0.028	-0.019	-0.039*
		(0.018)	(0.023)	(0.022)	(0.022)
	College graduate	0.258***	-0.035	-0.005	-0.018
		(0.020)	(0.025)	(0.024)	(0.024)
Race:	Black	-0.080***	-0.069**	-0.017	0.000
		(0.023)	(0.029)	(0.027)	(0.027)
	Asian	-0.111***	0.012	0.044	0.058
		(0.033)	(0.042)	(0.039)	(0.039)
	Other	0.047	-0.063	-0.099	-0.042
		(0.052)	(0.065)	(0.061)	(0.060)
Inc:	Below $20,000	-0.637***	0.030	0.043	0.046*
		(0.022)	(0.028)	(0.027)	(0.027)
	$20,000-39,999	-0.229***	-0.027	-0.020	-0.015
		(0.021)	(0.026)	(0.025)	(0.025)
	$70,000 and up	0.393***	0.022	0.030	0.030
		(0.021)	(0.025)	(0.024)	(0.024)
Constant		9.144***	0.111**	0.087*	0.118**
		(0.041)	(0.052)	(0.050)	(0.050)
		B: Adjusted means			
Total expenditures (log)		9.099***	9.145***	9.163***	9.181***
		(0.007)	(0.005)	(0.005)	(0.005)
		8,942***	9,369***	9,536***	9,711***
		(66)	(49)	(43)	(46)

Notes: This table replicates Table 3.7, but uses second-interview difficulty of reaching. $N = 36,722$ (1 attempt: 5,076; 2: 8,781; 3–4: 11,507; 5+: 11,358). $R^2 = 0.56$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.17: Total expenditures, with additional interview controls

Attempts		1	2	3–4	5+
		A: Regression with interactions			
		Base	Interactions		
Age:	16–29	-0.071***	-0.019	-0.007	0.006
		(0.024)	(0.034)	(0.031)	(0.029)
	30–39	-0.016	-0.002	0.001	-0.024
		(0.022)	(0.030)	(0.027)	(0.026)
	50–64	-0.017	-0.030	0.004	-0.013
		(0.020)	(0.027)	(0.025)	(0.024)
	65 and up	-0.056**	-0.027	-0.015	-0.038
		(0.022)	(0.030)	(0.029)	(0.028)
Children in household		0.055**	-0.043	-0.035	-0.021
		(0.022)	(0.030)	(0.028)	(0.027)
Female		-0.030**	0.019	0.006	0.021
		(0.012)	(0.017)	(0.016)	(0.015)
Educ:	Less than high school	-0.137***	0.025	0.018	0.002
		(0.019)	(0.026)	(0.025)	(0.024)
	Some college or tech. school	0.094***	-0.022	-0.024	-0.035*
		(0.016)	(0.022)	(0.020)	(0.019)
	College graduate	0.235***	-0.005	-0.023	-0.021
		(0.017)	(0.023)	(0.022)	(0.021)
Race:	Black	-0.065***	-0.014	-0.045*	-0.020
		(0.020)	(0.027)	(0.025)	(0.023)
	Asian	-0.049	-0.018	-0.040	0.011
		(0.031)	(0.041)	(0.038)	(0.036)
	Other	0.027	-0.093	-0.105*	-0.036
		(0.046)	(0.061)	(0.058)	(0.053)
Inc:	Below $20,000	-0.555***	-0.038	-0.014	0.006
		(0.019)	(0.027)	(0.025)	(0.024)
	$20,000-39,999	-0.233***	-0.014	0.005	0.013
		(0.017)	(0.024)	(0.023)	(0.022)
	$70,000 and up	0.384***	0.011	0.009	0.015
		(0.017)	(0.024)	(0.022)	(0.021)
Constant		9.014***	0.121**	0.148***	0.033
		(0.039)	(0.054)	(0.050)	(0.048)
		B: Adjusted means			
Total expenditures (log)		9.124***	9.143***	9.168***	9.165***
		(0.006)	(0.006)	(0.005)	(0.004)
Total expenditures ($)		9,175***	9,344***	9,586***	9,554***
		(58)	(53)	(46)	(42)

Notes: This table replicates Table 3.7, but adds additional controls (and number-of-attempts interactions) for interview hour (4 categories: 1–2pm, 3–4pm, 5pm or later) and duration (4 categories: less than 37.15, 37.15–52.78, 52.78–73.72, more than 73.72 minutes). 2 observations excluded due to missing interview duration. $N = 36,720$ (1 attempt: 6,603; 2: 7,229; 3–4: 9,971; 5+: 12,917). $R^2 = 0.59$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.18: Total expenditures, second-interview difficulty and additional interview controls

Attempts		1	2	3–4	5+
		A: Regression with interactions			
		Base	Interactions		
Age:	16–29	-0.125***	0.043	0.073**	0.053*
		(0.027)	(0.034)	(0.032)	(0.032)
	30–39	-0.005	-0.028	0.008	-0.047*
		(0.024)	(0.030)	(0.029)	(0.028)
	50–64	-0.040*	0.004	0.020	0.017
		(0.022)	(0.028)	(0.026)	(0.026)
	65 and up	-0.094***	0.013	0.025	0.010
		(0.025)	(0.031)	(0.030)	(0.030)
Children in household		0.020	0.019	-0.016	0.035
		(0.026)	(0.032)	(0.030)	(0.030)
Female		-0.022	0.015	0.005	-0.001
		(0.014)	(0.017)	(0.017)	(0.017)
Educ:	Less than high school	-0.102***	-0.054**	-0.023	-0.009
		(0.021)	(0.027)	(0.026)	(0.026)
	Some college or tech. school	0.106***	-0.042*	-0.032	-0.050**
		(0.018)	(0.023)	(0.022)	(0.022)
	College graduate	0.245***	-0.047*	-0.017	-0.027
		(0.020)	(0.024)	(0.023)	(0.023)
Race:	Black	-0.077***	-0.056*	-0.006	0.007
		(0.022)	(0.028)	(0.027)	(0.026)
	Asian	-0.084***	0.007	0.039	0.039
		(0.032)	(0.041)	(0.038)	(0.038)
	Other	0.018	-0.044	-0.089	-0.036
		(0.051)	(0.063)	(0.060)	(0.059)
Inc:	Below $20,000	-0.594***	0.028	0.037	0.035
		(0.022)	(0.027)	(0.026)	(0.026)
	$20,000-39,999	-0.210***	-0.027	-0.027	-0.021
		(0.020)	(0.025)	(0.024)	(0.024)
	$70,000 and up	0.377***	0.015	0.020	0.022
		(0.020)	(0.025)	(0.024)	(0.024)
Constant		8.993***	0.137**	0.102**	0.134**
		(0.042)	(0.054)	(0.052)	(0.052)
		B: Adjusted means			
Total expenditures (log)		9.126***	9.151***	9.158***	9.171***
		(0.007)	(0.005)	(0.004)	(0.005)
		9,187***	9,423***	9,490***	9,614***
		(69)	(49)	(42)	(45)

Notes: This table replicates Table 3.7, but uses second-interview difficulty of reaching. and adds additional controls (and number-of-attempts interactions) for interview hour (4 categories: 1–2pm, 3–4pm, 5pm or later) and duration (4 categories: less than 37.15, 37.15–52.78, 52.78–73.72, more than 73.72 minutes). 2 observations excluded due to missing interview duration. $N = 36,720$ (1 attempt: 5,075; 2: 8,781; 3–4: 11,506; 5+: 11,358). $R^2 = 0.59$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.19: Health expenditures

Attempts		1	2	3–4	5+
		A: Regression with interactions			
		Base	Interactions		
Age:	16–29	-1.140***	0.187	0.012	0.317**
		(0.123)	(0.171)	(0.156)	(0.145)
	30–39	-0.444***	0.040	-0.012	0.029
		(0.112)	(0.149)	(0.138)	(0.130)
	50–64	0.659***	-0.244*	-0.242*	-0.254**
		(0.102)	(0.135)	(0.126)	(0.120)
	65 and up	2.075***	-0.096	-0.144	-0.054
		(0.112)	(0.151)	(0.143)	(0.139)
Children in household		0.100	-0.193	-0.228	-0.109
		(0.112)	(0.153)	(0.142)	(0.134)
Female		0.280***	-0.001	0.071	0.053
		(0.061)	(0.085)	(0.079)	(0.075)
Educ:	Less than high school	-0.422***	-0.054	-0.155	-0.227*
		(0.095)	(0.133)	(0.126)	(0.121)
	Some college or tech. school	0.313***	-0.150	-0.091	-0.031
		(0.079)	(0.111)	(0.103)	(0.098)
	College graduate	0.579***	-0.159	0.013	0.033
		(0.086)	(0.118)	(0.110)	(0.105)
Race:	Black	-0.493***	0.007	-0.043	-0.072
		(0.100)	(0.138)	(0.127)	(0.118)
	Asian	-0.325**	-0.220	-0.258	-0.196
		(0.156)	(0.206)	(0.194)	(0.182)
	Other	-0.145	-0.463	0.280	-0.026
		(0.230)	(0.307)	(0.291)	(0.268)
Inc:	Below $20,000	-1.423***	-0.002	-0.015	-0.064
		(0.095)	(0.133)	(0.124)	(0.119)
	$20,000-39,999	-0.564***	-0.049	0.013	-0.093
		(0.087)	(0.122)	(0.114)	(0.109)
	$70,000 and up	0.403***	0.166	0.161	0.306***
		(0.088)	(0.119)	(0.111)	(0.106)
Constant		4.958***	0.249	0.104	-0.473**
		(0.183)	(0.252)	(0.235)	(0.225)
		B: Adjusted means			
Health expenditures (log)		5.080***	5.076***	5.048***	4.940***
		(0.032)	(0.028)	(0.024)	(0.022)
Health expenditures ($)		160***	159***	155***	139***
		(5)	(5)	(4)	(3)

Notes: This table replicates Table 3.7, but replaces the dependent variable with ln(total quarterly household expenditures on health +1) (and excludes households with negative net health expenditures). $N = 36,568$ (1 attempt: 6,571; 2: 7,199; 3–4: 9,926; 5+: 12,872). $R^2 = 0.27$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.20: Food expenditures

Attempts		1	2	3–4	5+
		A: Regression with interactions			
		Base	Interactions		
Age:	16–29	-0.092**	-0.037	-0.026	-0.035
		(0.036)	(0.050)	(0.046)	(0.043)
	30–39	-0.020	0.003	0.023	-0.052
		(0.033)	(0.044)	(0.041)	(0.038)
	50–64	-0.017	-0.020	-0.022	-0.018
		(0.030)	(0.040)	(0.037)	(0.035)
	65 and up	-0.087***	0.040	0.030	0.011
		(0.033)	(0.044)	(0.042)	(0.041)
Children in household		0.052	-0.007	-0.031	0.013
		(0.033)	(0.045)	(0.042)	(0.039)
Female		-0.060***	0.006	-0.010	0.004
		(0.018)	(0.025)	(0.023)	(0.022)
Educ:	Less than high school	-0.098***	0.063	0.046	0.058
		(0.028)	(0.039)	(0.037)	(0.035)
	Some college or tech. school	0.040*	0.003	0.003	-0.008
		(0.023)	(0.033)	(0.030)	(0.029)
	College graduate	0.150***	0.001	-0.004	-0.025
		(0.025)	(0.035)	(0.032)	(0.031)
Race:	Black	-0.045	-0.081**	-0.078**	-0.085**
		(0.030)	(0.041)	(0.037)	(0.035)
	Asian	-0.073	0.008	-0.017	0.034
		(0.046)	(0.060)	(0.057)	(0.053)
	Other	0.041	-0.063	-0.066	-0.102
		(0.068)	(0.090)	(0.086)	(0.079)
Inc:	Below $20,000	-0.358***	-0.030	-0.036	-0.057
		(0.028)	(0.039)	(0.037)	(0.035)
	$20,000-39,999	-0.146***	-0.023	-0.008	0.005
		(0.026)	(0.036)	(0.033)	(0.032)
	$70,000 and up	0.226***	0.013	0.006	0.021
		(0.026)	(0.035)	(0.032)	(0.031)
Constant		7.340***	0.007	0.125*	0.026
		(0.054)	(0.074)	(0.069)	(0.066)
		B: Adjusted means			
Food expenditures (log)		7.242***	7.253***	7.288***	7.301***
		(0.009)	(0.008)	(0.007)	(0.006)
Food expenditures ($)		1,397***	1,412***	1,462***	1,481***
		(13)	(12)	(10)	(10)

Notes: This table replicates Table 3.7, but replaces the dependent variable with ln(total quarterly household expenditures on food +1). $N = 36,724$ (1 attempt: 6,603; 2: 7,230; 3–4: 9,973; 5+: 12,918). $R^2 = 0.31$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.21: Total expenditures, telephone interviews

Attempts		1	2	3–4	5+
		A: Regression with interactions			
		Base	Interactions		
Age:	16–29	-0.090*	-0.063	-0.017	-0.013
		(0.053)	(0.071)	(0.064)	(0.059)
	30–39	-0.011	-0.010	-0.008	-0.027
		(0.046)	(0.060)	(0.055)	(0.051)
	50–64	-0.002	-0.055	-0.014	-0.026
		(0.042)	(0.054)	(0.049)	(0.046)
	65 and up	-0.025	-0.065	-0.053	-0.081
		(0.048)	(0.063)	(0.059)	(0.055)
Children in household		0.082*	-0.102	-0.088	-0.037
		(0.047)	(0.063)	(0.057)	(0.054)
Female		-0.022	0.021	0.004	0.030
		(0.026)	(0.035)	(0.032)	(0.030)
Educ:	Less than high school	-0.110**	0.015	-0.008	-0.032
		(0.046)	(0.062)	(0.057)	(0.054)
	Some college or tech. school	0.115***	-0.007	-0.039	-0.046
		(0.033)	(0.045)	(0.041)	(0.039)
	College graduate	0.272***	0.003	-0.025	-0.043
		(0.035)	(0.047)	(0.043)	(0.040)
Race:	Black	0.023	-0.152***	-0.163***	-0.142***
		(0.042)	(0.056)	(0.051)	(0.047)
	Asian	0.012	-0.049	-0.102	-0.093
		(0.058)	(0.076)	(0.071)	(0.066)
	Other	0.219*	-0.258*	-0.288**	-0.338***
		(0.115)	(0.153)	(0.135)	(0.125)
Inc:	Below $20,000	-0.577***	-0.066	0.002	0.042
		(0.041)	(0.056)	(0.051)	(0.048)
	$20,000-39,999	-0.215***	-0.011	0.016	-0.010
		(0.038)	(0.050)	(0.046)	(0.043)
	$70,000 and up	0.424***	-0.018	-0.012	0.016
		(0.036)	(0.047)	(0.043)	(0.041)
Constant		9.045***	0.141	0.243***	0.163*
		(0.076)	(0.101)	(0.093)	(0.088)
		B: Adjusted means			
Total expenditures (log)		9.143***	9.165***	9.208***	9.202***
		(0.013)	(0.011)	(0.009)	(0.007)
		9,344***	9,555***	9,972***	9,920***
		(124)	(107)	(87)	(71)

Notes: This table replicates Table 3.7, but restricts the sample to telephone interviews. $N = 11,719$ (1 attempt: 1,593; 2: 1,992; 3–4: 3,185; 5+: 4,949). $R^2 = 0.55$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.22: Total expenditures, personal interviews

Attempts		1	2	3–4	5+
		A: Regression with interactions			
		Base	Interactions		
Age:	16–29	-0.089***	-0.001	-0.000	0.019
		(0.029)	(0.040)	(0.037)	(0.035)
	30–39	-0.024	-0.000	0.003	-0.023
		(0.027)	(0.036)	(0.033)	(0.032)
	50–64	0.001	-0.022	0.003	-0.025
		(0.024)	(0.032)	(0.031)	(0.029)
	65 and up	-0.033	-0.016	-0.001	-0.020
		(0.026)	(0.036)	(0.034)	(0.034)
Children in household		0.054**	-0.024	-0.022	-0.016
		(0.026)	(0.036)	(0.034)	(0.032)
Female		-0.020	0.015	0.001	0.023
		(0.014)	(0.020)	(0.019)	(0.018)
Educ:	Less than high school	-0.166***	0.032	0.041	0.026
		(0.022)	(0.030)	(0.029)	(0.028)
	Some college or tech. school	0.095***	-0.016	-0.006	-0.018
		(0.019)	(0.026)	(0.025)	(0.024)
	College graduate	0.248***	-0.002	-0.012	-0.006
		(0.020)	(0.028)	(0.027)	(0.026)
Race:	Black	-0.109***	0.030	0.004	0.032
		(0.024)	(0.033)	(0.031)	(0.029)
	Asian	-0.115***	0.008	0.004	0.072
		(0.038)	(0.051)	(0.048)	(0.046)
	Other	0.022	-0.084	-0.078	0.064
		(0.052)	(0.069)	(0.067)	(0.062)
Inc:	Below $20,000	-0.624***	-0.009	0.010	0.015
		(0.022)	(0.031)	(0.030)	(0.029)
	$20,000-39,999	-0.270***	-0.012	0.011	0.028
		(0.020)	(0.029)	(0.027)	(0.026)
	$70,000 and up	0.387***	0.028	0.036	0.041
		(0.021)	(0.029)	(0.027)	(0.026)
Constant		9.218***	0.068	0.076	-0.031
		(0.043)	(0.060)	(0.056)	(0.055)
		B: Adjusted means			
Total expenditures (log)		9.097***	9.124***	9.144***	9.147***
		(0.007)	(0.007)	(0.006)	(0.006)
Total expenditures ($)		8,923***	9,170***	9,355***	9,387***
		(66)	(63)	(56)	(54)

Notes: This table replicates Table 3.7, but restricts the sample to personal interviews. $N = 24,305$ (1 attempt: 4,912; 2: 5,131; 3–4: 6,611; 5+: 7,651). $R^2 = 0.57$. *** p<0.01, ** p<0.05, * p<0.1.

Table B.23: Life satisfaction, uncontrolled OLS

	Separate				Cumulative		
Attempts:	1 call	2-3 calls	4-7 calls	≥8 calls	1–3 calls	1–7 calls	All
Inc. below $25,000	-0.317***	-0.319***	-0.313***	-0.301***	-0.318***	-0.316***	-0.312***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)
Inc. $25,000-49,999	-0.104***	-0.113***	-0.111***	-0.118***	-0.109***	-0.109***	-0.110***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Inc. $75,000 and up	0.110***	0.107***	0.115***	0.126***	0.108***	0.111***	0.114***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Constant	3.487***	3.487***	3.477***	3.464***	3.487***	3.483***	3.479***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.001)	(0.001)
R^2	0.056	0.057	0.059	0.060	0.056	0.057	0.058
Female	0.002	-0.002	-0.013***	-0.012***	-0.001	-0.005***	-0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Constant	3.369***	3.384***	3.401***	3.403***	3.377***	3.386***	3.390***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
R^2	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Age 18–39	0.042***	0.027***	0.013***	0.006**	0.033***	0.027***	0.023***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Age 40–49	-0.006	-0.006**	-0.004	-0.008**	-0.006**	-0.004**	-0.004**
	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Age 60–69	0.086***	0.085***	0.081***	0.076***	0.085***	0.082***	0.079***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)
Age 70 and up	0.082***	0.070***	0.058***	0.049***	0.074***	0.067***	0.061***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)
Constant	3.323***	3.347***	3.368***	3.380***	3.337***	3.348***	3.355***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
R^2	0.004	0.003	0.003	0.002	0.003	0.003	0.003
Age ≤ 49 & child in house	0.121***	0.107***	0.096***	0.080***	0.113***	0.106***	0.099***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Age ≥ 50	0.115***	0.107***	0.098***	0.084***	0.109***	0.103***	0.096***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)
Constant	3.268***	3.291***	3.312***	3.330***	3.281***	3.293***	3.303***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)
R^2	0.004	0.003	0.003	0.003	0.003	0.003	0.003
Constant	3.370***	3.383***	3.393***	3.396***	3.377***	3.382***	3.385***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	344,244	428,658	394,212	314,363	772,902	1,167,114	1,481,477

Notes: Source: Behavioral Risk Factor Surveillance System, 2005–2008. Each column reports coefficients from five different OLS regressions (separated by horizontal lines). Dependent variable: responses to the question "In general, how satisfied are you with your life?" coded as follows: 4=Very satisfied; 3=Satisfied; 2=Dissatisfied; 1=Very dissatisfied. Where relevant, regressions include indicators for missing data (not reported). Regressions within a column are conducted on the same subsample, based on the difficulty-to-reach category reported on the top row. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table B.24: Life satisfaction, controlled OLS

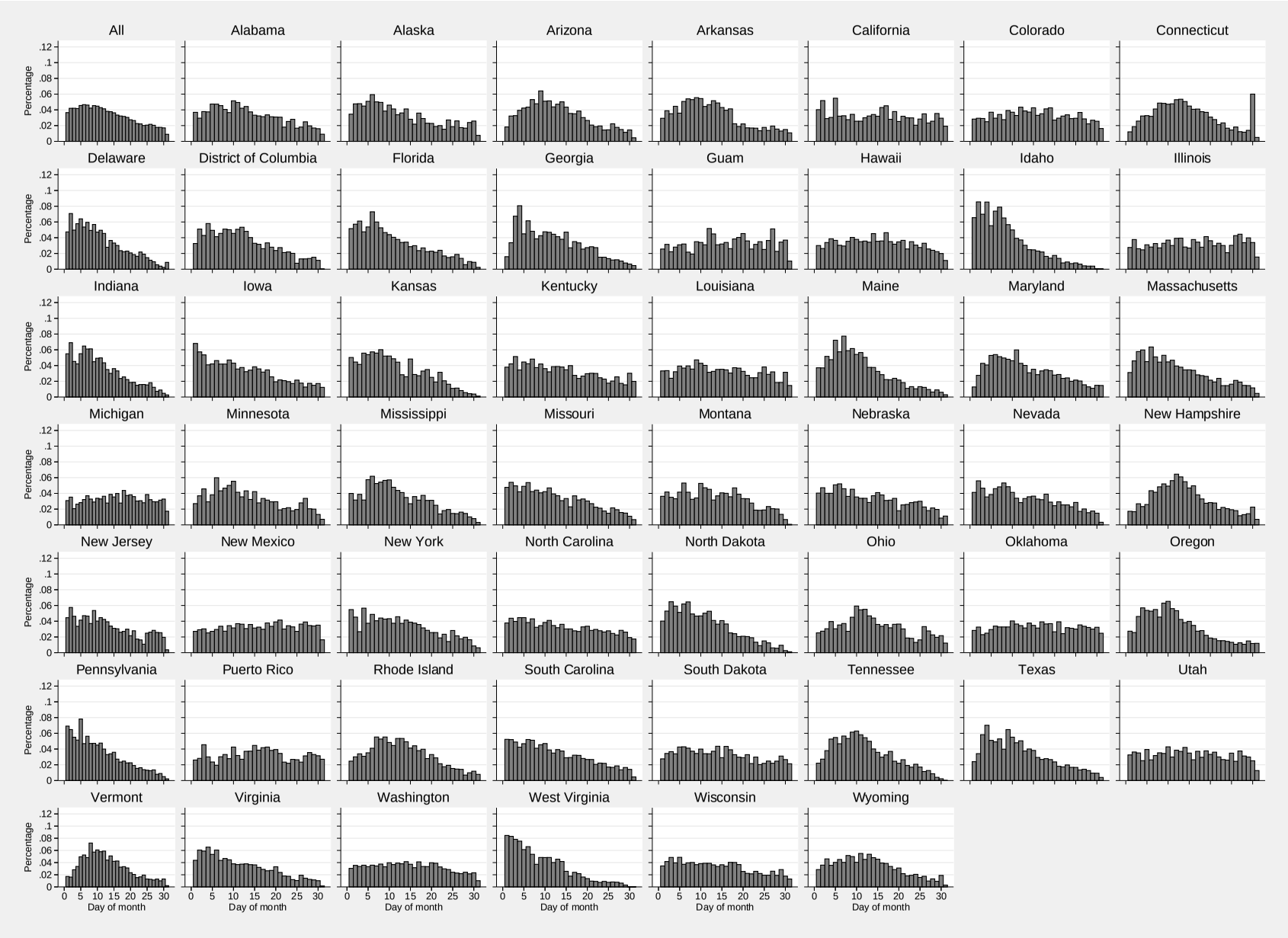
	Separate				Cumulative		
Attempts:	1 call	2-3 calls	4-7 calls	≥8 calls	1–3 calls	1–7 calls	All
Inc. below $25,000	-0.254***	-0.251***	-0.238***	-0.231***	-0.253***	-0.248***	-0.245***
	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)
Inc. $25,000-49,999	-0.090***	-0.095***	-0.086***	-0.088***	-0.093***	-0.090***	-0.090***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Inc. $75,000 and up	0.093***	0.088***	0.091***	0.096***	0.090***	0.091***	0.092***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Female	0.041***	0.040***	0.030***	0.029***	0.040***	0.036***	0.034***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Age 18–39	0.032***	0.038***	0.029***	0.033***	0.036***	0.034***	0.035***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.003)	(0.002)	(0.002)
Age 40–49	-0.035***	-0.028***	-0.025***	-0.020***	-0.030***	-0.028***	-0.025***
	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)
Age 60–69	0.128***	0.127***	0.122***	0.114***	0.127***	0.124***	0.121***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)
Age 70 and up	0.189***	0.184***	0.174***	0.162***	0.186***	0.181***	0.176***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)
Age ≤ 49 & child in house	0.025***	0.018***	0.020***	0.010***	0.021***	0.021***	0.018***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Constant	3.350***	3.335***	3.338***	3.388***	3.343***	3.342***	3.351***
	(0.011)	(0.010)	(0.011)	(0.013)	(0.007)	(0.006)	(0.005)
R^2	0.097	0.097	0.097	0.094	0.097	0.097	0.096
Observations	344,244	428,658	394,212	314,363	772,902	1,167,114	1,481,477

Notes: Source: Behavioral Risk Factor Surveillance System, 2005–2008. Each column reports coefficients from a single OLS regression. Dependent variable: responses to the question "In general, how satisfied are you with your life?" coded as follows: 4=Very satisfied; 3=Satisfied; 2=Dissatisfied; 1=Very dissatisfied. Regressions include controls for education, marital status, race, state, urban/rural, and missing data indicators that are not reported. The regression for each column is conducted on a subsample, based on the difficulty-to-reach category reported on the top row. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Figure B.1: Distribution of interview month, by location

Notes: Source: Behavioral Risk Factor Surveillance System, 2012. Sample: all interviews. Each histogram reports the distribution of the month of interview completion.

Figure B.2: Distribution of interview day of the month, by location

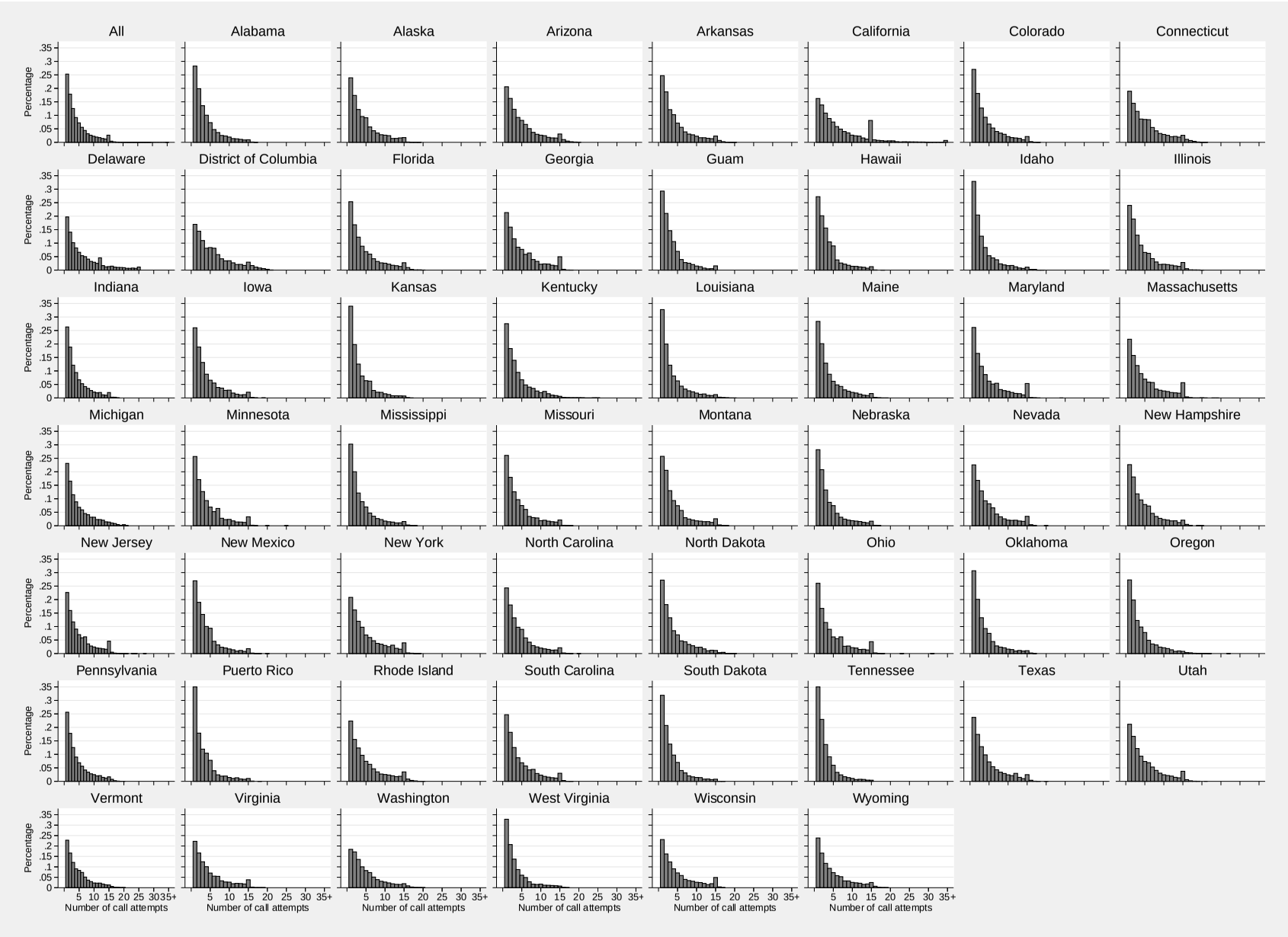


Notes: Source: Behavioral Risk Factor Surveillance System, 2012. Sample: all interviews. Each histogram reports the distribution of the day of the month of interview completion.

Figure B.3: Distribution of interview day of the week, by location

Notes: Source: Behavioral Risk Factor Surveillance System, 2012. Sample: all interviews. Each histogram reports the distribution of the day of the week of interview completion.

Figure B.4: Distribution of number of contact attempts, by location



Notes: Source: Behavioral Risk Factor Surveillance System, 2012. Sample: all interviews. Each histogram reports the distribution of the number of attempts required for each completion.

BIBLIOGRAPHY

AAPOR (2016). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*, 9th edition.

Bates, N., Dahlhamer, J., Phipps, P., Safir, A., and Tan, L. (2010). Assessing Contact History Paradata Quality Across Several Federal Surveys. In *JSM Proceedings of the Survey Research Methods Section of the American Statistical Association*, pages 91–105.

Behaghel, L., Crépon, B., Gurgand, M., and Le Barbanchon, T. (2014). Please Call Again: Correcting Non-Response Bias in Treatment Effect Models. *Review of Economics and Statistics*, 97(5):1070–1080.

Biemer, P., Chen, P., and Wang, K. (2011). Errors in the Recorded Number of Call Attempts and their Effect on Nonresponse Adjustments Using Callback Models. In *58th World Statistical Congress (Session IPS033)*.

Biemer, P. P., Chen, P., and Wang, K. (2013). Using Level-of-Effort Paradata in Non-response Adjustments with Application to Field Surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(1):147–168.

BRFSS (2013). 2012 Summary Data Quality Report. `http://www.cdc.gov/brfss/annual_data/2012/pdf/summarydataqualityreport2012_20130712.pdf` (accessed May 26, 2016).

Brick, J. M. and Williams, D. (2012). Explaining Rising Nonresponse Rates in Cross-sectional Surveys. *The ANNALS of the American Academy of Political and Social Science*, 645(1):36–59.

CDC (2012). Weighting the Data. `http://www.cdc.gov/brfss/annual_`

data/2012/pdf/Weighting-the-data_webpage-content-20130709.pdf (accessed May 26, 2016).

Chadi, A. (2014). Dissatisfied With Life or With Being Interviewed? Happiness and Motivation to Participate in a Survey. *SOEPpaper*, No. 639.

Chetty, R., Friedman, J. N., and Saez, E. (2013). Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings. *American Economic Review*, 103(7):2683–2721.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and Taxation: Theory and Evidence. *American Economic Review*, 99(4):1145–1177.

Cohen, S. B., Rohde, F., and Yu, W. (2013). Building Wave Response Rates in a Longitudinal Survey: Essential for Nonsampling Error Reduction or Last In-First Out? *Field Methods*, 25(4):361–387.

CPS (2015). Current Population Survey Interviewing Manual. http://www2.census.gov/programs-surveys/cps/methodology/intman/CPS_Manual_April2015.pdf (accessed May 26, 2016).

Curtin, R., Presser, S., and Singer, E. (2000). The Effects of Response Rate Changes on the Index of Consumer Sentiment. *Public Opinion Quarterly*, 64(4):413–428.

Diamond, P. (1980). Income taxation with fixed hours of work. *Journal of Public Economics*, 13(1):101–110.

Dixon, J. (2013). Using Paradata to Understand Panel Effects in the Current Population Survey Unemployment Rate. In *Proceedings of the 2013 Federal Committee on Statistical Methodology (FCSM) Research Conference*, pages 2–9.

Feldman, N., Katuscak, P., and Kawano, L. (2016). Taxpayer confusion: Evidence from the child tax credit. *American Economic Review*, 106(3):807–835.

Finkelstein, A. (2009). E-ZTax: Tax Salience and Tax Rates. *The Quarterly Journal of Economics*, 124:969–1010.

Gideon, M. (2017). Do individuals perceive income tax rates correctly? *Public Finance Review*, 45(1):97–117.

Glaeser, E. L., Gottlieb, J. D., and Ziv, O. (2016). Unhappy cities. *Journal of Labor Economics*, 34(S2):S129–S182.

Goldin, J. and Homonoff, T. (2013). Smoke Gets in Your Eyes: Cigarette Tax Salience and Regressivity. *American Economic Journal: Economic Policy*, 5:302–336.

Heffetz, O. and Rabin, M. (2013). Conclusions Regarding Cross-Group Differences in Happiness Depend on Difficulty of Reaching Respondents. *American Economic Review*, 103(7):3001–3021.

Hetschko, C. and Chadi, A. (2017). The Magic of the New: How Job Changes Affect Job Satisfaction. *Journal of Economics and Management Strategy*, forthcoming.

Keeter, S., Miller, C., Kohut, A., Groves, R. M., and Presser, S. (2000). Consequences of Reducing Nonresponse in a National Telephone Survey. *Public Opinion Quarterly*, 64(2):125–148.

Kreuter, F., Müller, G., and Trappmann, M. (2010). Nonresponse and Measurement Error in Employment Research: Making Use of Administrative Data. *Public Opinion Quarterly*, 74(5):880–906.

Krueger, A. B., Mas, A., and Niu, X. (2017). The Evolution of Rotation Group Bias: Will the Real Unemployment Rate Please Stand Up? *Review of Economics and Statistics*, 2(99):258–264.

Krueger, B. S. and West, B. T. (2014). Assessing the Potential of Paradata and Other Auxiliary Data for Nonresponse Adjustments. *Public Opinion Quarterly*, 78(4):795–831.

Legleye, S., Charrance, G., Razafindratsima, N., Bohet, A., Bajos, N., and Moreau, C. (2013). Improving Survey Participation: Cost Effectiveness of Callbacks to Refusals and Increased Call Attempts in a National Telephone Survey in France. *Public Opinion Quarterly*, 77(3):666–695.

Liebman, J. and Zeckhauser, R. (2004). Schmeduling.

Lin, I.-F. and Schaeffer, N. C. (1995). Using Survey Participants to Estimate the Impact of Nonparticipation. *Public Opinion Quarterly*, 59(2):236–58.

Meyer, B. D., Mok, W., and Sullivan, J. X. (2015). Household Surveys in Crisis. *Journal of Economic Perspectives*, 29(4):199–226.

Miller, B. and Mumford, K. (2015). The salience of complex tax changes: Evidence from the child and dependent care credit expxansion. *National Tax Journal*, 68(3):477–510.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies*, 38(2):175–208.

NRC (2013). Nonresponse in Social Science Surveys: A Research Agenda. Technical report, Panel on a Research Agenda for the Future of Social Science Data Collection, Committee on National Statistics. Division of Behavioral and Social Sciences and Education, Washington, DC.

Oswald, A. J. and Wu, S. (2010). Objective Confirmation of Subjective Measures of Human Well-Being: Evidence from the U.S.A. *Science*, 327(5965):576–579.

Potthoff, R. F., Manton, K. G., and Woodbury, M. A. (1993). Correcting for Nonavailability Bias in Surveys by Weighting Based on Number of Callbacks. *Journal of the American Statistical Association*, 88(424):1197–1207.

Pudney, S. and Watson, N. (2013). If at First You Don't Succeed? Fieldwork, Panel Attrition, and Health-Employment Inferences in BHPS. *ISER Working Paper Series*, 2013-27.

Rees-Jones, A. and Taubinsky, D. (2016). Heuristic perceptions of the income tax: Evidence and implications for debiasing.

Saez, E. (2010). Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy*, 2(3):180–212.

Taubinksy, D. and Rees-Jones, A. (2017). Attention variation and welfare: Theory and evidence from a tax salience experiment. *Review of Economic Studies (forthcoming)*.

WHO (1995). The Use and Interpretation of Anthropometry Physical Status: Report of a WHO Expert Committee. Technical report, WHO Techincal Report Series, No. 854, Geneva.

WHO (2000). Obesity: Preventing and Managing the Global Epidemic: Report of a WHO Consultation. Technical report, WHO Techincal Report Series, No. 894, Geneva.

Williams, R. (2011). Using the Margins Command to Estimate and Interpret Adjusted Predictions and Marginal Effects. In *Stata Conference, Chicago*.