American Economic Journal: Economic Policy 2015, 7(1): 275–299
http://dx.doi.org/10.1257/pol.20120081

Would People Behave Differently If They Better Understood Social Security? Evidence from a Field Experiment†

By Jeffrey B. Liebman and Erzo F. P. Luttmer*

This paper presents the results of a randomized field experiment that provided information about key Social Security features to older workers. The experiment was designed to examine whether it is possible to affect individual behavior using a relatively inexpensive informational intervention about the provisions of a public program and to explore the mechanisms underlying the behavior change. We find that our relatively mild intervention (sending an informational brochure and an invitation to a web-tutorial) increased labor force participation one year later by 4 percentage points relative to the control group mean of 74 percent. (*JEL* C93, D12, H55)

The provisions of government tax, social insurance, and means-tested transfer programs create complex sets of incentives for individuals making labor supply, retirement, and savings decisions. If individuals do not understand or do not otherwise come to correctly perceive the incentives, they may make economic decisions that are privately suboptimal and may also fail to participate effectively as political actors.[1]

An important policy question is whether there exist relatively inexpensive approaches to providing information that improve decision making and ultimately increase individual well-being. For example, to what extent could a simple brochure mailed by the Social Security Administration (SSA) correct the widespread misperceptions about the Social Security earnings test? The answer to this question will depend in part on why incentives are currently misperceived. In some cases, the necessary information may be straightforward to understand, but expensive (in either monetary or psychic terms) to acquire. In other cases, the information about program rules may be readily available, but the calculation necessary to determine an individual's own incentives may be very complicated. In still other cases, cognitive biases may cause people

* Liebman: John F. Kennedy School of Government, Harvard University, 79 JFK Street, Cambridge, MA 02138 (e-mail: jeffrey_liebman@harvard.edu); Luttmer: Department of Economics, Dartmouth College, 6106 Rockefeller Center, Hanover, NH 03755 (e-mail: erzo.fp.luttmer@dartmouth.edu). We thank Michael Anderson, Dan Benjamin, John Geanakoplos, Jeffrey Kling, David Laibson, Annamaria Lusardi, Brigitte Madrian, Susann Rohwedder, Mark Shepard, Stephen Zeldes, several anonymous referees, and seminar participants at Collegio Carlo Alberto, Harvard University, the NBER Summer Institute, and Stanford University for helpful comments. We thank Jonathan Li, Kate Mikels, Abdul Tariq, and Victoria Levin for superb research assistance. We thank Michael Anderson for sharing Stata code on family-wise error rates. This research was supported by the U.S. Social Security Administration through grant #10-M-98363-1-01 to the National Bureau of Economic Research as part of the SSA Retirement Research Consortium. The findings and conclusions expressed are solely those of the authors and do not represent the views of SSA, any agency of the Federal Government, or the NBER. All errors are our own.
† Go to http://dx.doi.org/10.1257/pol.20120081 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.
[1] In some cases, these privately suboptimal decisions can be socially optimal. See Liebman and Zeckhauser (2008).

to misperceive even relatively simple incentive schedules. Finally, powerful social cues may point people toward a suboptimal decision, even when the correct information is also readily available. While factors such as these that can produce poor decision making are well documented, there is little evidence on how easy they are to overcome.

This paper presents the results of a field experiment in which a sample of older workers was randomized between a treatment group that was given information about key Social Security provisions, and a control group that was not given the information. One year after the information was provided, we administered a follow-up survey and measured the impact of the information provision on labor supply and Social Security benefit claiming behavior. We find that our relatively mild intervention (a mailed brochure combined with an invitation to participate in a 15-minute online tutorial) raised the fraction of the sample remaining in the labor force by 4 percentage points. This impact is statistically significant for the entire sample and appears to be driven by female sample members, who increased their labor force participation by 7 percentage points in response to the treatment. The difference between females and males in the size of their response to the treatment is not, however, statistically significant. We do not find statistically significant effects of the treatment on the probability that respondents have started claiming Social Security benefits.

Because our intervention provided treatment group members with several different pieces of information about Social Security and also communicated a more general message that "working additional years is beneficial," it is impossible to isolate which aspects of the intervention led to the labor force participation response. Nonetheless, to explore the mechanisms by which the intervention might have affected behavior, we ask respondents a series of questions about their understanding of the incentives that the Social Security program provides for labor supply and benefit claiming. We find that the information intervention increased the perceived return to working longer, especially among women, which is consistent with a pathway in which the intervention affected behavior by changing knowledge of incentives. However, we do not have evidence that would allow us to determine whether this pathway was the only mechanism, or even the primary mechanism, through which the impact occurred.

I. Background

It is becoming increasingly clear that responses of economic actors to the incentives created by government tax and spending programs are affected not only by the size of the incentives, but also by contextual factors that affect how the incentives are perceived.[2] Duflo et al. (2006) show that customers of a tax preparation firm who were offered a match of contributions to retirement savings accounts were much more responsive to the simple and transparent match offer than are US taxpayers who face a similar match via a provision in the US tax code. Chetty, Looney, and Kroft (2009) find that changes in excise taxes yield larger behavioral responses than economically equivalent changes in sales taxes, most likely because the sales taxes are added at the tax register and are therefore less salient. Kling et al. (2012) posit

[2] Bernheim and Rangel (2009) refer to these contextual factors as ancillary conditions.

that many Medicare beneficiaries are poorly informed about the prices offered by different prescription drug plans and show that providing information can cause beneficiaries to switch to lower-priced plans. Chetty and Saez (2013) demonstrate that EITC recipients alter their labor supply when the incentives provided by the program are explained by a tax preparer, and Song (2013) shows that contribution levels to a highly subsidized government pension scheme increase by 40 percent among subjects who receive a financial education treatment that emphasizes interest compounding. Mastrobuoni (2011) shows that the mailing of Social Security statements by the SSA increases knowledge about benefit levels. While Mastrobuoni estimates small and statistically insignificant impacts on retirement behavior, the 95 percent confidence interval around these estimates is wide and ranges from a 108 percent increase in the impact of Social Security incentives on retirement to a 66 percent decrease.

The context-specific nature of behavioral responses to incentives increases the dimensionality of the challenge for researchers seeking to reach a consensus about the magnitude of behavioral responses to policy provisions, because there will not be a single parameter that can be averaged across multiple studies if the studies measure behavior in disparate contexts. However, the sensitivity of behavioral responses to how incentives are perceived by individual decision makers also provides policymakers with an additional instrument; relatively inexpensive interventions that provide information or alter the framing of decisions have the potential to significantly improve economic well-being.[3]

Social Security policy is likely to be a particularly fruitful area in which to apply these insights. Decisions about when to retire and when to claim benefits can have large implications for well-being over many subsequent years. Such decisions also have elements of irreversibility that make it hard to undo poor decisions. And retirement-related decisions are very challenging to get right.[4] The Social Security tax and benefit schedules themselves involve several features that make it hard to perceive incentives correctly—complex nonlinearities and interactions with other sources of retirement income that make calculations difficult, a remote connection between choices and payoffs because benefits are often not received until many years after the point at which labor supply decisions are made, and a dependence of one's own incentives on individual-specific factors, such as marital history and life expectancy, that make it hard to learn one's own incentives from that of a peer. In addition to schedule complexity, these choices involve uncertainty and tradeoffs over time—the two contexts in which difficulty in decision making have been the most widely documented.[5] Simulations by Benítez-Silva, Demiralp, and Liu (2009) using a dynamic life-cycle model suggest the potential for large welfare gains from improving

[3] In a recent field experiment, Goda, Manchester, and Sojourner (2012) find that a low-cost, direct-mail intervention affects contributions to defined-contribution retirement accounts.

[4] There is a large literature on the accuracy of people's perceptions about future levels of Social Security and pension benefits and the implications of lack of knowledge on retirement and saving behavior. See Bernheim (1988); Mitchell (1988); Gustman and Steinmeier (2005a); Rohwedder and Kleinjans (2006); Dominitz, Hung, and van Soest (2007); and Chan and Huff Stevens (2008). See also Lusardi and Mitchell (2009) for an analysis of the link between financial literacy and retirement planning.

[5] See Liebman and Zeckhauser (2004) for a discussion of decision making under complex schedules. See Beshears et al. (2008) and Liebman and Zeckhauser (2008) for a review of the literature on contexts in which individuals have difficulties making wise decisions.

people's understanding of Social Security. Moreover, experiments have shown effects of how information about Social Security is presented on hypothetical or intended behavior—raising the possibility that interventions could also alter actual behavior.[6]

How people perceive the incentives of Social Security factors critically in several policy debates. For example, there is no consensus on the extent to which people correctly perceive the marginal future Social Security benefits they receive when they work an additional hour.[7] Knowing the answer to this question is important for understanding the amount of deadweight loss caused by the OASDI payroll tax and also for assessing the potential welfare gains from switching to a system—either based on notional defined-contribution accounts or funded personal retirement accounts—with more transparent linkage between initial collections and later benefits. Similarly, there remains no consensus about why so many people retire and claim benefits at age 62.[8] Knowing the answer to this question is important for understanding the welfare implications for adjusting the earliest eligibility age and the full-benefit age.

In an earlier paper (Liebman and Luttmer 2012), we administered a survey about Social Security benefit rules to a representative sample of Americans aged 50–70. We found that the majority of respondents believe that their Social Security benefits increase with an additional year of work, i.e., that the Social Security benefit rules provide a positive work incentive. The magnitude of the perceived incentive varies across respondents, but people generally cite an incentive that is somewhat larger than our best prediction of this incentive. However, our predictors of the true incentives are not sufficiently precise to rule out moderate over- or under-perceptions of incentives in the general population or in population subgroups. We also surveyed people about their understanding of various provisions in the Social Security benefit rules, and found that some of these provisions (e.g., effects of delayed benefit claiming and rules on widow benefits) are relatively well understood while others (rules on spousal benefits and provisions on which years of earnings are taken into account) are less well understood.

In order to achieve the potential welfare gains from improving people's understanding of the tradeoffs they face, researchers need to make progress in two areas. The first is in understanding the reasons for current misperceptions. The second is in testing interventions to discover which approaches are most effective in correcting misperceptions. Our current study makes three contributions in these areas. First, it provides further confirmation that interventions affecting people's perceptions of incentives can affect individual behavior in important policy contexts. Second, by measuring both the change in behavior *and* the change in people's perceptions, it provides suggestive evidence regarding how the intervention may have worked to change behavior. Third, the study demonstrates an approach to altering perceived incentives—a combination of a mailing and an online tutorial—that is both

[6] Hastings and Tejeda-Ashton (2008) show in the context of the privatized Mexican Social Security system that financially illiterate workers are more price sensitive in their hypothetical choice of retirement fund when fees are expressed in pesos rather than in terms of annual percentage rates. Brown, Kapteyn, and Mitchell (forthcoming) and Liebman and Luttmer (2012) find that the intended age of claiming Social Security benefits depends on how the early versus late claiming decision is framed.

[7] See Liebman, Luttmer, and Seif (2009) for evidence on this subject.

[8] Gustman and Steinmeier (2005b) attribute the spike in retirement at age 62 to high discount rates. Cutler et al. (2013) present evidence that focal point behavior is the main cause of the spike.

relatively inexpensive and which allows full researcher control over the information that is provided. In particular, it does not rely on a caseworker-style approach in which one-on-one counseling is used to influence people. For many applications, the approach we demonstrate is likely to be scalable to the population level once an effective intervention is identified. Our approach also removes the ambiguity over what information or message was provided to the sample member. This ambiguity, along with cost, is a drawback of the caseworker approach.

II. Survey Design and Experimental Manipulations

A. *Population for the Information Intervention*

We contracted with Knowledge Networks to administer our experimental intervention and a follow-up survey to a sample drawn from its panel of respondents. These panelists, originally recruited through random-digit dialing, agree to take a 15–20 minute survey once a week via the Internet using a PC or WebTV in exchange for free Internet or WebTV access.[9] The panelists receive incentive payments and rewards through a loyalty program. Knowledge Networks collects basic demographic characteristics for all its panelists, and its panelists are roughly representative of the adult US population according to these characteristics.

In 2008, we fielded a baseline survey of knowledge about Social Security using the Knowledge Network panel (for details see Liebman and Luttmer 2012). The sample for that survey spanned the ages of 30 to 70, with an oversample of working individuals between the ages of 60 and 65. For the current study, we started with the 3,611 individuals who were invited to participate in our 2008 survey. We dropped 668 individuals who were younger than 55 years old, 419 individuals who were not working (based on the employment status variable that is part of the Knowledge Networks standard demographic profile variables), and 41 individuals who told us in the 2008 survey that they were not covered by Social Security. This procedure resulted in an experimental sample of 2,483 working individuals who were randomized between the treatment group and the control group. Most of our sample members (90 percent) were between the ages of 60 and 65 in November 2008. We focus on working individuals near the retirement age because that is the population most likely to display behavioral effects of the information intervention within the time frame of our experiment.

B. *Information Intervention*

The members of the treatment group received an informational intervention that consisted of two parts. In February of 2009, Knowledge Networks sent them by regular mail an informational brochure that we had created for this purpose. In March of 2009, members of the treatment group were invited to participate in a web-tutorial about Social Security. The completion rate of the web-tutorial was 76.8 percent for the treatment group as a whole and 91.6 percent for treatment group members who

[9] The WebTV option means that individuals did not need to be computer users to be recruited into the panel.

completed the follow-up survey.[10] Members of the control group did not receive any materials from us (either in the mail or online).

In both the brochure and the web-tutorial, we provided information on three topics. First, we provided information about longevity to emphasize the need for an adequate planning horizon for retirement finances. For example, the brochure mentions, "For the average 65-year-old couple living in America today, there is a 47 percent chance that at least one spouse will live to age 90." Second, we explained the relationship between retirement age and the standard of living during retirement. In particular, we explain how Social Security benefits rise with the age at which benefits are claimed and how Social Security benefits depend on a person's work history. For example, the web-tutorial introduces a hypothetical worker and shows what happens to this person's Social Security benefits if the person works more years. Third, we provide information about the Social Security earnings test and explain that cuts in current benefits due to the earnings test are offset by higher benefits in the future. In designing the information intervention, we took into account that many respondents may have trouble remembering benefit formulae and information presented with figures and statistics. We therefore complemented such information with vignettes about actual retirees whom we had interviewed and from whom we had received permission to incorporate their information in our materials. These vignettes helped to underline the broader message of the intervention and contribute to the "gestalt" of the message.

Before giving information about Social Security in the brochure, we explained that the brochure was sent to them by Knowledge Networks on behalf of us ("researchers at Harvard University") as a follow-up to a recent survey that the individual had been invited to take (our 2008 baseline survey). Moreover, we emphasized that the brochure was not a comprehensive source of Social Security facts, and we provided phone numbers and links to additional sources of information about Social Security. We printed the brochure on glossy paper, used a relatively large font, and provided it with a professional layout in color in order to entice respondents to read it. The complete brochure is included in online Appendix A.

The web-tutorial covered the same three basic topics as the brochure, but was tailored to each respondent's situation—something we were able to do because it was administered online. For example, we gave information about the typical longevity of people of the same sex as the respondent; the characters in examples had the same sex as the respondent, and returns to delaying claiming were calculated exactly based on the respondent's birth cohort. The web-tutorial also contained a number of questions about the information we presented in order to induce the respondents to pay attention to the information. In the web-tutorial, as in the mailed brochure, we emphasized the gestalt of the information intervention by including vignettes of actual retirees. For example, the tutorial contained the following vignette:

"Among those nearing retirement age, it's common to worry about future finances, and for good reason: retirees must grapple with the reality of paying for living expenses for years to come. However, with a bit of planning, retirees can live comfortably well beyond retirement age. The following is a story about 91-year-old

[10] As we explain below, some sample members were no longer active with Knowledge Networks at the time the invitations to participate in the web-tutorial were sent out.

Leon, who says, 'I'm the other side of 91.' [Picture of Leon is shown] Leon worked as a chemistry teacher for 26 years before retiring at age 70. Each year, he put aside the maximum amount out of his paycheck toward his pension. He also knew that waiting until age 70 to claim Social Security would increase his monthly benefits. Leon says that his current financial situation is 'better than I've ever had. Between my pension and Social Security...I don't worry about anything.' According to Leon, people approaching retirement should continue to work as long as they're healthy."

At the end of the web-tutorial, we provided phone numbers and Internet links to further resources for information about Social Security. The complete web-tutorial is included in online Appendix B.[11]

C. *Follow-Up Survey*

In April 2010, 13 months after the information intervention, members from both the treatment group and the control group were invited to participate in our follow-up survey. The follow-up survey was designed to measure the effects of the information intervention on labor supply and Social Security benefit claiming behavior. In order to explore what mechanisms underlie the behavior change, it also included questions intended to measure understanding of the incentives that the Social Security program provides for labor supply and benefit claiming. In addition, because one year is a relatively short period in which to observe changes in retirement behavior, the survey also contained questions about planned future behavior. The follow-up survey contained 67 questions and the median time to complete it was 18 minutes. We paid respondents a $5 incentive for completing the survey. Inactive panelists were offered an additional $5 to return to the panel to take the survey.[12] The follow-up survey was fielded between April 8 and June 9 of 2010. While the vast majority (88.5 percent) of respondents completed the survey in April, we kept the survey open until June to maximize the response rate. The full survey instrument is provided in online Appendix C.

Of the 2,483 members of the experimental sample, 1,596 completed the follow-up survey for an overall response rate of 64.3 percent. Online Appendix Figure A1 contains a flow chart describing the evolution of our analysis sample. It shows that attrition from the experimental sample occurred in three steps and that most of it appears to have been for reasons unrelated to the intervention. Online Appendix Figure A2 shows a timeline for the different stages of data collection.

First, only 89 percent of experimental sample members were invited to take the follow-up survey because 11.4 percent of the sample had permanently left the Knowledge Networks panel or had informed Knowledge Networks that they were temporarily unavailable for surveys. The attrition rate at this first step was similar between the treatment group (11.0 percent) and control group (11.7 percent).

[11] There is little overlap between the information provided in our survey and the information provided in the Social Security statement that was being mailed annually to workers at the time of our survey. The 35-year rule is not mentioned in the Social Security statement, and the statement says only that future benefits "could" increase if you are affected by the earnings test.

[12] Panelists can become "inactive" either by voluntarily withdrawing from the panel or by being retired from the regular panel by Knowledge Networks. However, Knowledge Networks retains the possibility of inviting inactive members for particular surveys, e.g., for surveys that involve a longitudinal dimension.

Second, not all sample members who were invited to take the survey clicked on the link to learn the topic of the survey. In particular, 78.3 percent of invited treatment group members and 74.2 percent of invited control group members clicked on the link. The particular way in which invitations to take the follow-up survey were extended minimized the opportunity for treatment-induced differential non-response between the treatment and control groups. As is typical of Knowledge Network survey invitations, the invitation simply invited sample members to take a survey—without specifying the subject of the survey—and provided a link to click on if the sample member wanted to participate. Moreover, the overwhelming majority (86 percent) of invited sample members who failed to click on the link had been identified by Knowledge Networks as inactive panelists prior to the date of our survey invitation. Thus over 90 percent of the combined attrition that occurred in these first and second steps occurred because of respondents who were no longer actively participating in Knowledge Networks surveys, rather than from a decision to skip our particular survey.

Third, conditional on clicking on the link and thereby learning the topic of the survey, 3.8 percent of treatment group members and 4.0 percent of control group members failed to complete the survey.[13] Conditional on completing the survey, the item-response rates were very high, generally well above 95 percent.

In forming our analysis sample, we drop one observation of a sample member for whom the age according the Knowledge Network profile variable increased by three years between November 2008 and April 2010, which is logically impossible. This yields a final sample of 1,595 observations for our main analyses.

D. *Analysis of Attrition Patterns*

Despite the blinded nature of the survey invitation, the overall response rate to the follow-up survey was 4.4 percentage points higher in the treatment group than the control group, and this difference is significant at the 5 percent level. As described above, this difference can be explained by the fact that members of the treatment group were more likely to still be active Knowledge Networks panelists at the April first date of the survey invitation. Among panelists active as of the invitation date, the response rates for the treatment group and for the control group were not significantly different from each other. And nearly all of the differential attrition occurred before sample members were aware of the survey topic. Thus, the differential response rate is unrelated to the topic or content of the follow-up survey.

The concern, of course, is that the attrition process may not be random and may result in treatment and control populations that would no longer have the same expected outcomes if they received the same treatment. To explore this issue, we examined whether the demographic characteristics of the treatment and control groups differ among takers of the follow-up survey. These results are presented in online Appendix Table A2. Panel A tests whether these demographic characteristics are jointly statistically different between the two groups by regressing treatment

[13] An additional 1 percent of respondents in each group was skipped out of the survey early in administration once screening questions revealed that they were not eligible for Social Security.

status on the full set of demographics listed in panel B. Consistent with successful random assignment and no differential attribution by demographics, the full set of demographics is jointly insignificant (p-value 0.759). Panel B shows that the samples are well balanced on the individual demographic variables as well. Only 1 of 35 pairwise comparisons has a p-value below 0.05. We, nevertheless, control for baseline demographics in our regressions in order to increase the statistical precision of treatment effect estimates and to adjust for any random differences between the treatment and control groups in observable characteristics.

Even though the treatment and control groups are balanced on observable demographics, it is possible that they differ on some unobserved characteristics. This would be particularly worrisome if something about the treatment caused treatment groups members to be more willing to take the survey in a way that directly affected the mean of the outcome variable. For example, if treatment group members who were working were more inclined to answer the survey because they were proud to exhibit behavior that was consistent with the general message of the intervention, then our results could be biased toward finding a positive impact of the intervention on labor supply. However, as we noted above, the differential attrition occurred before respondents were aware of the survey topic, so we can rule out this sort of mechanism.[14]

With direct channels between survey participation and treatment group/outcome status ruled out, it is very unlikely that a 4 percent difference in response rates could bias the results in a meaningful way. Indeed, we can bound the potential bias. Under the extreme assumption that all of the extra response in the treatment group was made up of people who were employed, this differential response would account for less than half of the estimated overall employment impact and less than one quarter of the impact among women.[15]

In summary, while we cannot completely rule out bias due to the differential attrition, there is no evidence in the data to suggest lack of balance between the treatment and control groups, the blinded nature of the survey invitation rules out some of the most troublesome types of attrition bias, and the magnitude of the differential attrition is small enough that it would take a very extreme attrition process to significantly alter the findings of this paper.[16]

[14] In theory, the information intervention could have affected panelist decisions about whether to remain active with Knowledge Networks. Given that panelists take weekly surveys, it is highly unlikely that our survey (out of the 50 or so that a typical panelist would have taken during a year) would have had a quantitatively significant impact on the decision to remain active. It is also conceivable that by affecting the probability of work, the intervention affected the availability of time to take surveys. However, this mechanism would lead to lower participation by those in the treatment group, which is the opposite of what we find.

[15] These calculations are performed as follows. The control group mean employment rate was 74.4 percent (Table 3). The control group response rate was 62.0 percent and the treatment group response rate was 66.4 percent (the differential response was 4.4 percentage points). If there were no true treatment effect but all the extra response in the treatment group consisted of working individuals, then the treatment group mean employment rate would have been $(0.620 \times 0.744 + 0.044 \times 1.00)/0.664 = 0.761$ or an impact of 1.7 percent relative to 74.4 percent. The upper bound of 1.7 percent is well below half of the 4.2 percent overall employment estimate.

[16] The differential attrition rates were similar among males (4 percent) and females (5 percent). If differential attrition were responsible for a significant portion of the estimated impact of the intervention, we would expect to see employment rate effects for both males and females. Instead, the employment effects are concentrated among females.

Table 1—Summary Statistics

	Women			Men			p-value on t-test of difference
	N	Mean	SD	N	Mean	SD	
Key outcome variables							
Did paid work last month (2010)	880	0.75	0.43	712	0.79	0.41	0.088
Hours worked for pay last month (2010)	879	91	76	712	109	79	0.000
Earnings last month (2010), in dollars/month	873	2,888	4,998	703	4,248	5,777	0.000
Not receiving Social Security benefits in 2010 (Asked if not claiming in 2008 and aged 60+ in 2010)	686	0.70	0.46	587	0.72	0.45	0.261
Demographics measured at the time of the baseline survey (2008)							
Age in 2010	883	63.2	2.2	712	63.1	2.2	0.419
Non-Hispanic white	883	0.88	0.32	712	0.90	0.30	0.257
Non-Hispanic black	883	0.06	0.23	712	0.03	0.18	0.034
Other race/ethnicity	883	0.06	0.24	712	0.07	0.25	0.749
High school dropout	883	0.02	0.13	712	0.01	0.11	0.328
High school	883	0.18	0.39	712	0.10	0.30	0.000
Some college	883	0.33	0.47	712	0.32	0.47	0.582
Bachelor's degree or more	883	0.47	0.50	712	0.57	0.50	0.000
Married	883	0.55	0.50	712	0.78	0.41	0.000
Widowed	883	0.07	0.26	712	0.02	0.14	0.000
Divorced	883	0.27	0.44	712	0.11	0.31	0.000
Separated	883	0.01	0.09	712	0.01	0.09	0.892
Never married	883	0.06	0.25	712	0.05	0.22	0.344
Living apart	883	0.03	0.18	712	0.03	0.18	0.977
Lives in the Northeast	883	0.20	0.40	712	0.17	0.37	0.137
Lives in the Midwest	883	0.26	0.44	712	0.25	0.43	0.511
Lives in the South	883	0.28	0.45	712	0.32	0.47	0.078
Lives in the West	883	0.26	0.44	712	0.26	0.44	0.913
Household size of one	883	0.35	0.48	712	0.21	0.41	0.000
Household size of two	883	0.49	0.50	712	0.57	0.50	0.001
Household size of three or more	883	0.16	0.37	712	0.22	0.41	0.005
Household income: less than 25k	883	0.09	0.28	712	0.05	0.21	0.002
Household income: 25k–50k	883	0.25	0.44	712	0.15	0.35	0.000
Household income: 50k–75k	883	0.25	0.43	712	0.22	0.42	0.167
Household income: 75k–100k	883	0.21	0.41	712	0.19	0.40	0.437
Household income: 100k or more	883	0.20	0.40	712	0.39	0.49	0.000

Notes: Key outcome variables are measured in the April–June 2010 Social Security Follow-Up Survey, designed by the authors and fielded by Knowledge Networks. The baseline demographics are the values in the standard demographic profile variables at the time of the baseline survey (November 2008). The standard demographic profile is collected by Knowledge Networks. Our sample is restricted to individuals working in November 2008 and heavily oversamples individuals between the ages of 60 and 65 as of November 2008.

III. Results

A. *Sample Characteristics*

Table 1 shows the demographic composition of the sample that completed the follow-up survey. We split the table by the gender of the respondent because labor market responses often differ by gender. About three quarters of the sample performed paid work in the calendar month previous to the follow-up survey, and this fraction is only slightly lower for women. The labor force participation rate may seem high for this age group, but recall that only individuals who were working in 2008 were eligible to participate in the experiment. Overall, about 40 percent of sample members were receiving Social Security benefits in 2010, which implies that a non-negligible fraction was combining work and benefit receipt. Of those who did not receive benefits in 2008 but were potentially eligible to receive Social Security benefits in 2010, 30 percent of women and 28 percent of men were receiving benefits.

We present the demographic characteristics of sample members at the time of the baseline survey (2008) because the regressions control for demographics as measured in 2008. We include the 2008 controls rather than the 2010 controls in the regressions because the treatment could conceivably affect some of the control variables in 2010. The specific controls that we present and use in regressions are maintained by Knowledge Networks and are therefore available for all sample members, even those who did not participate in the 2008 survey. Knowledge Networks recruits its panelists such that the demographic characteristics of its panelists are broadly representative of those in the US adult population. However, because we conditioned our sample on the respondent being in a narrow age range (90 percent of our panel is between 60 and 65 in November 2008) and on the respondent working in 2008, the demographic characteristics of our sample are not representative of the general population but instead are broadly representative of working individuals approaching the age at which most people are making retirement decisions.[17]

The mean age in our sample in 2010 is 63. About 89 percent of the sample is white, and about 84 percent has at least some college education. These percentages would be high for the general population in our age group, but recall that our sample is limited to working individuals. About 55 percent of the female respondents are married as are about 78 percent of the men. This difference is mostly accounted for by the fact that working women in this age range are much more likely to be divorced than are working men. The sample is geographically dispersed, with all regions of the country well represented. About 80 percent of respondents live in one- or two-person households. Finally, respondents come from households throughout the income distribution, though higher-income households are disproportionately represented because of the higher-than-average labor market participation in our sample.

B. *Manipulation Checks*

To verify that our experimental manipulation was effective and was implemented correctly, we asked two questions about the information intervention to all respondents at the end of the follow-up survey. We first asked them whether they recalled receiving "about a year ago . . . the following informational brochure" about Social Security rules, showing them a picture of the front page of our informational brochure on the screen. We then asked them whether they recalled participating "about

[17] Online Appendix Table A1 examines the representativeness of the experimental sample by comparing its demographic characteristics to those of observations in the Current Population Survey (CPS) that match our sample selection criteria of working and being between the ages of 60 and 65. Relative to the CPS, our experimental sample has somewhat more whites (85 percent versus 81 percent), is more educated (47 percent with a bachelor's versus 37 percent), is slightly less likely to be married (64 percent versus 70 percent), and has somewhat lower incomes (46 percent with annual incomes exceeding $75,000 versus 51 percent). While many of the demographic characteristics are statistically significantly different between the CPS sample and the experimental sample, the economic magnitude of these differences is moderate in size. We therefore think of the sample as broadly representative of the US population of working individuals between the ages of 60 and 65. In terms of demographic characteristics, the 1,595 respondents to the follow-up survey look very similar to the 2,483 individuals of the experimental sample, with the exception that respondents to the follow-up survey are 3.8 percentage points more likely to be white and 4.5 percentage points more likely to have a bachelor's degree. Overall, our survey sample is quite similar to our full experimental sample and, as we discussed earlier, there is no statistically significant demographic difference between the treatment group and the control group in the survey sample.

Table 2—Manipulation Checks

	Entire sample		Female respondents		Male respondents		
Dependent variable	Control mean [N of reg.]	Treatment effect (SE)	Control mean [N of reg.]	Treatment effect (SE)	Control mean [N of reg.]	Treatment effect (SE)	p-value on t-test of difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. Recall receiving our brochure about SS	0.048 [1,595]	0.286*** (0.018)	0.040 [883]	0.307*** (0.024)	0.059 [712]	0.256*** (0.027)	0.761
2. Recall taking our web-tutorial on SS	0.010 [1,595]	0.084*** (0.011)	0.012 [883]	0.083*** (0.015)	0.009 [712]	0.082*** (0.016)	0.993

Notes: Robust standard errors between parentheses. Number of observations in the regression sample in square brackets. Treatment effects are estimated by an OLS regression with controls for age and age squared as well as demographics measured at the time of the baseline survey. The demographic control variables, measured at the time of the baseline survey, consist of gender, race/ethnicity dummies (non-hispanic white, non-hispanic black, other race/ethnicity), education dummies (high school dropout, high school graduate, some college, college plus), marital status dummies (married, widowed, divorced, separated, never married, living apart), regional dummies (Northeast, Midwest, South, West), and household size dummies (one-person, two-person, 3+ person household). See online Appendix C for the exact wording of the questions that define the outcome variables: Q8.4 and Q8.5. Column 7 reports the p-value on the test of the hypothesis that treatment effects are equal for females and males.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

a year ago … in an online module that provided additional information about Social Security rules, and which was tailored to each person's individual information." We further told the respondents, "To help you remember this survey from the many other surveys you have taken, the online module contained stories about the experiences of two retirees, 91-year-old Leon and 66-year-old Elena," and showed them the same pictures of these two retirees as they had seen in the online module.

The first row of Table 2 presents the results for the recall rate of the brochure. The first column shows that 4.8 percent of sample members in the control group report receiving the brochure, even though they should not have received, and to the best of our knowledge did not receive, the brochure. We believe this 4.8 percent may have confused our brochure with other mailings they have received about Social Security. The second column shows that the recall rate is 28.6 percentage points higher for respondents in the treatment group than for those in the control group, and that this effect is highly significant.[18] For consistency with the other regressions in the paper, the treatment effect is estimated in an OLS regression of the outcome variable (recall of the brochure) on an indicator of belonging to the treatment group and a set of demographic controls, including a quadratic in age. The estimate is extremely similar if the demographic controls are omitted. The significant treatment effect on the brochure recall rate is reassuring in that it shows that our treatment had an impact on the respondents and appears to have been implemented correctly. Yet, even in the treatment group, the recall rate is only 33.4 percent. In other words, a majority of respondents in the treatment group do not recall receiving the brochure, even though they should all have received it. Because it is possible to be influenced

[18] Our finding that a significant fraction of sample members recall a mailing about Social Security is consistent with the Mastrobuoni (2011) finding that the annual mailing of Social Security benefit statements affects expectations of future benefit levels.

by information without recalling where or when one received the information, we do not believe the treatment effects on other outcome variables should be scaled up by the reciprocal of the treatment effect on recall. We therefore do not treat the recall rate as a first-stage regression for estimating a treatment-on-the-treated effect.

Columns 3 and 4 show the control group recall rate and the treatment effect for female respondents while columns 5 and 6 show these estimates for male respondents. Although the treatment effect is slightly larger for female respondents than for male respondents, this difference is not statistically significant, as column 7 indicates.

The second row of Table 2 shows the results for the recall rate of participating in the web tutorial on Social Security. While only 1.0 percent incorrectly recalls taking part in this tutorial, the treatment effect is only 8.4 percentage points. Though this estimate is highly statistically significant, it implies that merely 9.4 percent of respondents in the treatment group recall participating in the web tutorial. We know that 77 percent of individuals in the treatment group in fact participated in the web tutorial, and even among the participants the recall rate is only 10.2 percent. We surmise that the low treatment effect on recall of the web tutorial is related to the fact that the respondents take online surveys from Knowledge Networks quite frequently (typically a couple per month), and it is hard for them to recall with confidence based on the relatively limited information we provided them whether they took our specific web tutorial. It is also possible that because this question was asked at the end of the survey, respondents may have feared follow-up questions if they answered "yes." The remaining columns of row 2 show that the treatment effect on recall of the web tutorial is very similar for women and men.

C. *Impacts of the Intervention on Labor Supply and Benefit Claiming Behavior*

The experiment was designed to investigate the effect of better knowledge about the Social Security benefit rules on (i) labor supply and (ii) Social Security claiming behavior. The first three rows of panel A of Table 3 present the effects on labor supply; the final row of panel A contains the results for claiming behavior.

Our simplest measure of labor supply is the answer to the question whether the respondent performed any paid work in the previous calendar month (generally March 2010). In the control group, 74.4 percent of respondents worked in the previous month. This percentage may seem surprisingly high at first, but recall again that only individuals who were working in 2008 were included in the experiment. The information intervention increased this percentage by 4.2 percentage points, and this effect is significant at the 5 percent level. The estimate of the treatment effect is especially large in the female subsample. Female respondents are 7.2 percentage points more likely to work if they received the information intervention, while male respondents in the treatment group are a statistically insignificant 0.3 percentage points more likely to work. However, the male-female difference is not statistically significant (p-value of 0.102). Hence, we cannot rule out that the labor supply response to our intervention was the same for women and men.

We also measured labor supply by hours worked in the previous calendar month and own earnings in the previous calendar month. Consistent with our finding that labor force participation increased, we find positive point estimates for the effect of

Table 3—Treatment Effects on Behavior

Dependent variable	Entire sample: Control mean [N of reg.] (1)	Entire sample: Treatment effect (SE) (2)	Female respondents: Control mean [N of reg.] (3)	Female respondents: Treatment effect (SE) (4)	Male respondents: Control mean [N of reg.] (5)	Male respondents: Treatment effect (SE) (6)	p-value on t-test of difference (7)
Panel A. Primary outcome measures							
1. Did paid work last month (2010)	0.744 [1,592]	0.042** (0.021)	0.713 [880]	0.072** (0.029)	0.783 [712]	0.003 (0.031)	0.102
2. Hours worked for pay last month (2010)	97.7 [1,591]	2.8 (3.8)	88.6 [879]	4.8 (5.1)	108.9 [712]	1.4 (5.9)	0.659
3. Earnings last month (2010), in dollars/month	3,258 [1,576]	388 (263)	2,554 [873]	683** (336)	4,146 [703]	168 (443)	0.354
4. Not receiving Social Security benefits in 2010 (asked if not claiming in '08 and aged 60+ in '10)	0.719 [1,273]	−0.006 (0.023)	0.715 [686]	−0.032 (0.032)	0.724 [587]	0.023 (0.032)	0.227
Panel B. Indices of outcome measures							
5. Index of labor market outcomes (standardized average of standardized values of (1), (2), and (3))	1.000 [1,592]	0.082* (0.049)	1.000 [880]	0.154** (0.068)	1.000 [712]	0.017 (0.075)	0.174
6. Index of all key outcomes (standardized average of standardized values of (1), (2), (3), and (4))	1.000 [1,595]	0.064 (0.048)	1.000 [883]	0.119* (0.066)	1.000 [712]	0.072 (0.073)	0.637

Notes: Robust standard errors between parentheses. Number of observations in the regression sample in square brackets. The entire sample consists of 1,595 observations (883 females, 712 males), but some regressions may have smaller samples because of item nonresponse. The sample size for row 4 is 1,273 rather than 1,595 because we excluded 322 observations that were already claiming benefits in 2008 or were younger than 60 in 2010. These observations were excluded because the information treatment could not have possibly affected their claim status in 2010. The indices in rows 5 and 6 are standardized by the control group mean and standard deviation of the sample listed in the column. Treatment effects are estimated by an OLS regression with controls for age and age squared as well as demographics measured at the time of the baseline survey. See the note to Table 2 for a description of these demographic control variables. See online Appendix C for the exact wording of the questions that define the four outcome variables: Q1.13, Q1.16, Q1.17, and Q1.2, respectively. Column 7 reports the p-value on the test of the hypothesis that treatment effects are equal for females and males.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

the information intervention on hours worked and earnings, but only the estimate on earnings for female respondents is significant at the 5 percent level. The dispersion across respondents in hours worked and earnings is relatively high, which increases standard errors relative to the mean value of the variable, and which makes it more difficult to detect a statistically significant effect. For example, the treatment effect for the participation variable needs to exceed only 5.6 percent of the control group mean in order for it to be statistically significant. For the hours and earnings variables, however, it needs to reach at least 7.7 percent and 15.8 percent of the control group mean, respectively, to attain statistical significance.

It is not the case that the information intervention primarily induced labor force participation at only very minimal hours or earnings. If we redefine labor force participation to include just individuals who work at least 20 hours per month, we continue to find significant effects of the information treatment on labor force participation in the entire sample and in the subsample of female respondents. If we redefine labor force participation to include only individuals with at least $500 in monthly earnings, we continue to find significant effects of the information treatment on female labor force participation, though the effect for the entire sample is not statistically significant (p-value 0.110).

The final row of panel A of Table 3 shows results regarding the claiming of Social Security benefits. The sample is smaller than that for the labor supply results because we measured benefit claiming only among those who were 60 and older but not claiming benefits at the time of our baseline survey (November 2008). The table shows that just over 70 percent of those who were 60 or older but not claiming benefits at the time of our baseline survey have not yet started claiming Social Security benefits by the time of the follow-up survey (April/May 2010). The information treatment did not have a significant effect on this percentage.

The presence of multiple outcome measures increases the risk that some of them are statistically significant by chance even if in truth there is no treatment effect. We address this multiple inference problem in two ways. First, we reduce the number of outcome variables by creating summary indices, an approach that originated in biostatistics (O'Brien 1984). Second, we calculate p-values that are adjusted for the multiple inference problem using the Westfall and Young (1993) free stepdown resampling algorithm. Both approaches are increasingly used in economics (e.g., Kling, Liebman, and Katz 2007; Acemoglu and Finkelstein 2008; Anderson 2008; and Heckman et al. 2010).

In panel B of Table 3, we present indices of the primary outcome measures. Because claiming behavior is conceptually different from labor supply, we created two indices: one of the three labor supply measures, and one that combines the labor supply measures with claiming behavior. We construct the indices by first standardizing each of the primary outcome variables by the control group mean and variance, then taking a simple average of the non-missing values of the standardized variables, and finally standardizing the average by the control group mean and variance. Thus, the index measures effect sizes in terms of standard deviations of the control group.

Row 5 presents the index of labor supply outcomes, and shows that the information treatment increased labor supply in the entire sample by 8.2 percent of a control-group standard deviation, but that this effect is only marginally statistically significant. As before, the labor supply effect is most pronounced in the female subsample, where the effect is larger in magnitude (15.4 percent of a control-group standard deviation) and statistically significant. Thus, when we eliminate the multiple inference problem of having three labor supply outcome measures by creating a single labor supply index, we still find a significant labor supply response to the treatment among females. In row 6, we also include Social Security claiming as a component of the index. Not surprisingly, this weakens the results because we found no treatment effects on claiming behavior, but we still find a marginally significant effect on the index of all four outcome variables for females (p-value 0.071).

The second way of accounting for the multiple inference problem is to calculate the probability that at least one hypothesis out of a family of hypotheses is falsely rejected. This probability is the family-wise error rate (FWER). We calculate the FWER using the Westfall and Young (1993) free stepdown resampling algorithm.[19]

[19] We use 100,000 replications.

When the family of hypotheses consists of all four primary outcome measures equaling zero, only the treatment effect on female labor force participation is statistically significant (FWER-adjusted p-value of 0.046). This implies that the treatment effect on female labor force participation remains significant even after accounting for the fact that we tested for a treatment effect in four outcome measures. Alternatively, one can control for the family-wise error rate by considering a family of two hypotheses: no treatment effect on the index of the three labor supply measures and no treatment effect on claiming behavior. In this case, the treatment effect on the index of labor supply measures remains significant for females (FWER-adjusted p-value of 0.045). Overall, we conclude from panel B of Table 3 and the calculations of the FWER-adjusted p-values that the significant treatment effect on female labor supply holds up after accounting for the multiple inference problem.

Table 4 examines the robustness of the estimates in Table 3 to two alternative specifications: (i) the omission of control variables and (ii) a probit regression for binary outcome variables and a median regression for continuous outcome variables. Table 4 shows that the key findings of Table 3 are robust. The treatment effect on female labor force participation remains statistically significant and similar in magnitude for all specifications. Similar to the findings in Table 3, the estimated effects on hours and earnings are positive and consistent in magnitude with the increase in labor force participation, but only occasionally statistically significant. Also in line with the findings of Table 3, we never detect a significant effect of our treatment on Social Security claiming behavior.

IV. Understanding the Results

This section contains additional analysis and discussion of the results, with a focus on three questions: (i) what aspects of the intervention led to the behavioral responses; (ii) why do the experimental impacts appear to be concentrated among female sample members; and (iii) whether we would expect to see additional behavioral responses if we were able to measure outcomes after more time has passed.

A. *Which Aspects of the Intervention Led to the Behavioral Response?*

In designing this project, we and others we consulted with had significant doubts about whether the relatively mild intervention that was feasible given our resources could affect understanding about Social Security and alter behavior.[20] We were particularly concerned that if we tested too weak an intervention and found no impact, we would have learned little—since it would always be possible that a slightly stronger intervention would have had an impact. We therefore decided to combine several different approaches to providing information in order to maximize the strength of the intervention. Specifically, we offered each treatment-group sample member both an informational mailing and an online tutorial. Within each intervention, we

[20] Lusardi and Mitchell (2007) similarly caution that one should not expect meaningful impacts from one-time financial literacy interventions—not because the financial education is ineffective per se but because the "cure" is likely to be inadequate.

Table 4—Robustness of Treatment Effects on Behavior

	Entire sample		Female respondents		Male respondents		
Dependent variable	Control mean [N of reg.] (1)	Treatment effect (SE) (2)	Control mean [N of reg.] (3)	Treatment effect (SE) (4)	Control mean [N of reg.] (5)	Treatment effect (SE) (6)	p-value on t-test of difference (7)
Panel A. Did paid work last month (2010)							
1. Baseline	0.744	0.042**	0.713	0.072**	0.783	0.003	0.102
(row 1 of Table 3)	[1,592]	(0.021)	[880]	(0.029)	[712]	(0.031)	
2. No controls	0.744	0.040*	0.713	0.069**	0.783	0.004	0.124
	[1,592]	(0.021)	[880]	(0.029)	[712]	(0.031)	
3. Probit	0.744	0.041*	0.713	0.073**	0.783	0.001	0.089*
	[1,592]	(0.021)	[880]	(0.030)	[712]	(0.031)	
Panel B. Hours worked for pay last month (2010)							
4. Baseline	97.7	2.8	88.6	4.8	108.9	1.4	0.659
(row 2 of Table 3)	[1,591]	(3.8)	[879]	(5.1)	[712]	(5.9)	
5. No controls	97.7	3.0	88.6	4.6	108.9	0.9	0.629
	[1,591]	(3.9)	[879]	(5.1)	[712]	(5.9)	
6. Median regression	97.7	0.0	88.6	20.0	108.9	0.0	0.325
	[1,591]	(10.6)	[879]	(14.9)	[712]	(13.8)	
Panel C. Earnings last month (2010), in dollars/month							
7. Baseline	3,258	388	2,554	683**	4,146	168	0.354
(row 3 of Table 3)	[1,576]	(263)	[873]	(336)	[703]	(443)	
8. No controls	3,258	454*	2,554	646*	4,146	193	0.412
	[1,576]	(271)	[873]	(336)	[703]	(437)	
9. Median regression	3,258	500**	2,554	375	4,146	0	0.446
	[1,576]	(226)	[873]	(249)	[703]	(424)	
Panel D. Not receiving Social Security benefits in 2010 (asked if not claiming in 2008 and 60 or older in 2010)							
10. Baseline	0.719	−0.006	0.715	−0.032	0.724	0.023	0.227
(row 4 of Table 3)	[1,273]	(0.023)	[686]	(0.032)	[587]	(0.032)	
11. No controls	0.719	−0.019	0.715	−0.037	0.724	0.001	0.461
	[1,273]	(0.025)	[686]	(0.035)	[587]	(0.037)	
12. Probit	0.719	−0.011	0.715	−0.037	0.724	0.020	0.280
	[1,273]	(0.027)	[686]	(0.037)	[587]	(0.038)	
13. Including observations under 60 or that were already claiming SS in 2008	0.619	0.001	0.601	−0.019	0.642	0.032	0.219
	[1,595]	(0.021)	[883]	(0.028)	[712]	(0.031)	

Notes: Robust standard errors between parentheses. Number of observations in the regression sample in square brackets. The entire sample consists of 1,595 observations (883 females, 712 males), but some regressions may have smaller samples because of item nonresponse. The sample size for rows 10–12 is 1,273 rather than 1,595 because we excluded 322 observations that were already claiming benefits in 2008 or were younger than 60 in 2010. These observations were excluded because the information treatment could not have possibly affected their claim status in 2010. Treatment effects are estimated by regressions that control for age and age squared as well as demographics measured at the time of the baseline survey. See the note to Table 2 for a description of these demographic control variables. The baseline specification is an OLS regression. The Probit specification reports marginal effects. See online Appendix C for the exact wording of the questions that define the outcome variables: Q1.13, Q1.16, Q1.17, and Q1.2, respectively. Column 7 reports the p-value on the test of the hypothesis that treatment effects are equal for females and males.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

combined specific information about Social Security rules, more general information about the importance of taking steps to ensure adequate income during retirement, and vignettes in which actual retirees discussed their own experiences in ways that reinforced the more specific information.

The multi-faceted treatment raises the question of which aspects of the intervention led to the behavioral responses. Knowing the answer to this question is relevant if one were thinking of scaling up this intervention or of applying this finding to another policy domain. Would a mailing alone be sufficient? Was the specific knowledge about the program essential, or are vignettes the best way to communicate information in a salient manner? Are there interaction effects from presenting information in multiple ways?

Knowing the answers to these questions is important not only for learning how to scale up interventions at the least cost, but also for drawing normative policy recommendations from the findings. In particular, if the behavioral impact comes about because people now understand the program better and therefore make better decisions, then the normative implications are clear—we have made sample members better off by providing them with information. In contrast, if the intervention had its effect because it communicated an overall message of "continuing to work until older ages has benefits," then the normative implications are more complex; such an intervention could persuade people to continue to work for whom it is not optimal to do so.[21]

While our multifaceted intervention does not provide us with the ability to isolate the specific mechanisms that produced the labor supply impacts, we did ask questions to explore whether the intervention changed the amount of knowledge sample members had about different aspects of Social Security. Specifically, we examined the effects on knowledge about three main topics on which the information intervention focused: the likelihood that retirees live into their 90s, the effects of working longer and claiming later on Social Security benefits, and the earnings test.

Here we summarize the main findings. In the interest of space, we focus on the statistically significant findings in Table 5, but the full set of results is available in online Appendix Table A3. Further details are available in Liebman and Luttmer (2011).

The intervention did not significantly affect knowledge about longevity or the earnings test. It also did not significantly affect knowledge about Social Security benefits increasing with years worked or the percent increase in benefits per additional year worked. However, when we asked about work incentives in a more intuitive way ("Do you get a better deal or worse deal from Social Security if you work more years?"), we find that the intervention raised the fraction of the sample reporting that one gets a better deal from working more years by 4.3 percentage points from a base of 36.3 percent (p-value: 0.082). This effect is especially pronounced for female respondents, for whom the information treatment increases the fraction perceiving

[21] Because our information delivery mechanism, unlike the caseworker approach, allows for complete researcher control over what information is provided, it would in theory be straightforward to do follow-up experiments to determine which components of the intervention are necessary to produce the behavioral response we observed. One would simply randomize people into different treatments each containing different subsets of the experimental intervention. In practice, one would have to be selective since sample size considerations would limit the number of different permutations that could be tested.

Table 5—Selected Treatment Effects on Knowledge about Social Security and on Expected Future Behavior

	Entire sample		Female respondents		Male respondents		
Dependent variable	Control mean [*N* of reg.] (1)	Treatment effect (SE) (2)	Control mean [*N* of reg.] (3)	Treatment effect (SE) (4)	Control mean [*N* of reg.] (5)	Treatment effect (SE) (6)	*p*-value on *t*-test of difference (7)
Panel A. Knowledge about incentives for working more years							
1. SS a better deal if working more years	0.363 [1,536]	0.043* (0.025)	0.346 [848]	0.079** (0.034)	0.385 [688]	0.009 (0.038)	0.167
2. Aware that SS benefits are based on some number of years with the highest earnings	0.364 [1,528]	0.055** (0.025)	0.378 [842]	0.074** (0.034)	0.347 [686]	0.037 (0.037)	0.458
Panel B. Knowledge about incentives for claiming later							
3. SS increases for a typical worker for delaying claiming between ages 62 and 66	0.906 [1,572]	0.012 (0.014)	0.888 [873]	0.020 (0.021)	0.928 [699]	0.004 (0.020)	0.566
4. SS increases for a typical worker for delaying claiming between ages 66 and 70	0.803 [1,572]	0.023 (0.019)	0.788 [873]	−0.005 (0.028)	0.823 [699]	0.056** (0.026)	0.112
5. SS remains the same for a typical worker for delaying claiming between ages 70 and 74	0.367 [1,572]	−0.013 (0.024)	0.360 [873]	−0.012 (0.033)	0.375 [699]	−0.012 (0.037)	0.999
Panel C. Expected future claiming behavior							
6. R's point estimate of the expected or realized SS claim age (if not claiming in 2008)	65.5 [1,307]	0.07 (0.13)	65.6 [698]	−0.15 (0.20)	65.4 [609]	0.33* (0.18)	0.069*
7. The mean of R's pdf of the expected or realized SS claim age (if not claiming in 2008)	65.5 [1,203]	0.08 (0.11)	65.5 [647]	−0.11 (0.15)	65.5 [556]	0.34** (0.15)	0.037**

Notes: Robust standard errors between parentheses. Number of observations in the regression sample in square brackets. The entire sample consists of 1,595 observations (883 females, 712 males), but some regressions may have smaller samples because of item nonresponse. Rows 6 and 7 have a smaller sample because the question was only asked of the relevant subsample of respondents. Treatment effects are estimated by an OLS regression with controls for age and age squared as well as demographics measured at the time of the baseline survey. See the note to Table 2 for a description of these demographic control variables. See online Appendix C for the exact wording of the questions that define the outcome variables. Panel A: Q3.3 combined with Q3.1 and Q5.1, respectively. Panel B: Q4.2. Panel C: Q1.8 and Q1.33, respectively. Column 7 reports the *p*-value on the test of the hypothesis that treatment effects are equal for females and males. Percentage increases in Social Security per year of delay in claiming benefits are measured as a percentage of benefits at age 66, and are top and bottom coded at +/− 25 percent. Note that this table only reports selected treatment effects. The complete set of treatment effects on knowledge variables are in online Appendix Table A3 and the complete set of treatment effects on planned or expected outcomes are in online Appendix Table A4.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

a better deal by 7.9 percentage points from a base of 34.6 percent (*p*-value: 0.019), though we cannot reject that the effect is the same for female and male respondents. This pattern dovetails with the pattern that we found for labor supply responses: much larger point estimates on the female labor supply response to our intervention than on the male labor supply response, although we cannot reject that the response is the same for females and males.[22]

We also ask respondents a multiple choice question about which years of earnings determine one's Social Security benefits. In line with results in our earlier paper, just under 40 percent of individuals in the control group are aware that the benefits

[22] An alternative explanation of why our intervention did not change perceptions of the increase in benefits from additional years of work but did change perceptions about Social Security being a good deal could be that it increased perceived life expectancy. However, we do not find any impact of the intervention on the perception of life expectancy.

are based on their X highest years of earnings.[23] The second row of Table 5 shows that the information intervention significantly raises the fraction of respondents that answers that benefits are based on years with highest earnings by 5.5 percentage points. Again, the point estimates are substantially larger for female respondents, for whom the effect is 7.4 percentage points and statistically significant, although we cannot reject the hypothesis that the effect for men and women is the same.

When we examine knowledge about incentives for claiming benefits at a later date, we find that over 90 percent of respondents know that benefits increase when a worker delays claiming between the ages of 62 and 66, and over 80 percent know that benefits increase when claiming is delayed between the ages of 66 and 70.[24] The mean perceived increase in benefits per year of delayed claiming as a percentage of benefits at age 66 is quite accurate for the 62–66 age range (perception of 7.1 percent versus actual value of 6.25 percent), is too low for the 66–70 age range (perception of 3.1 percent versus actual value of 8.0 percent), and too high for the 70–74 age range (perception of 3.4 percent versus actual value of 0 percent). The information treatment had no significant effects on these responses for the sample as a whole, though it increased the fraction of male respondents that perceive a positive return for delaying claiming between the ages of 66 and 70 by a statistically significant 5.6 percentage points. Similarly, the information treatment has a positive and statistically significant effect on male respondents' perceptions of the percentage increase in benefits per year of delayed claiming in the same 66-to-70 age range (see online Appendix Table A3).[25]

Table 5 and online Appendix Table A3 show that the information intervention did not lead to large and significant shifts in average responses for many of our knowledge questions. We therefore want to be careful not to overinterpret the significant effects that we do find. Nevertheless, we believe that, taken as a whole, the estimates of treatment effects on knowledge suggest three points. First, to the extent we find significant effects, these effects all indicate an increased awareness of the benefits of working longer and claiming later. This finding suggests that at least part of the behavioral labor supply response may have occurred through a higher perceived return to working more years. Second, the treatment had an especially large and significant effect on the perceived incentive for women to work more years, which matches our earlier finding that the labor supply response was especially strong among female respondents. Third, the information intervention has a significant effect on men's perceptions of the return to delaying claiming between the ages of 66 and 70, which, as we will see in Section IVB below, dovetails with the increase in planned Social Security claim ages among male respondents. While these results are suggestive of a pathway in which

[23] While 36 percent of controls are aware that Social Security benefits are based on the years with the highest earnings, very few know that the correct number of years is 35. The median answer is 9 years and only 7 percent give the correct answer of 35.

[24] See Liebman and Luttmer (2011) for details on how these variables are measured.

[25] It is interesting to speculate about why our intervention did not affect actual Social Security claiming behavior given that claiming behavior is arguably the place where people's decisions most deviate from what the standard economic model says they should do. It may simply be that teaching people that Social Security is a valuable real annuity and that they should buy more of it by claiming later is a more difficult concept to communicate than "you may live a long time, so don't stop working too early." It would be interesting to design a follow-up intervention that focused more specifically on educating people about this issue.

the intervention affected behavior by changing perceptions of incentives, we cannot rule out the possibility that the "gestalt" of the information intervention was partly or fully responsible for the observed responses in realized and planned behavior.

B. *Treatment Effects on Planned or Expected Behavior*

Because little more than one year elapsed between the administration of the information intervention and the collection of the follow-up data, the time span during which respondents could possibly change their behavior in response to the new information was limited.[26] In an attempt to capture effects of the information intervention on future behavior, we also asked the respondents a number of questions about their planned future behavior. Specifically, we asked them about their expected retirement age, likelihood of continuing to work after starting to collect Social Security benefits, and expected date of claiming Social Security benefits. We think of these responses as more speculative than our responses on current behavior because the planned behavior measures are likely to be less reliable (cheap talk) and more noisy (respondents might not yet have firm plans or have thought future decisions through) than realized behavior.

Overall, the information intervention had no significant effects on planned or expected outcomes. We suspect that this lack of statistically significant findings is partly due to the fact that expectations are not measured as precisely as realized outcomes. The one exception to the lack of statistically significant findings is that the information intervention increased the expected claim age among male respondents, as shown in panel C of Table 5. While we do not want to give too much credence to one significant finding out of multiple outcome variables, we note that this change in expected behavior matches our earlier finding that the information intervention increased the perceived return to delaying claiming among men. We take this correspondence between changes in perceived incentives and changes in expected behavior as an indication that the intervention may have worked at least partly through changing perceptions of incentives. The full results on planned and expected outcomes are discussed in Liebman and Luttmer (2011) and available in online Appendix Table A4.

C. *Why Do Behavioral Impacts Seem to Be Concentrated among Females?*

The labor supply impacts in Table 3 are especially pronounced for female sample members, raising the question of why this is so. However, despite the intriguing difference in the point estimates between the female and male labor supply response, we noted that we cannot reject that the responses are the same in magnitude. Given that the difference in labor supply responses is not even statistically significant itself, we generally do not have sufficient statistical power to convincingly determine the mechanism behind this difference because explorations into the mechanism involve examining subsamples of our data or further interaction terms. Hence,

[26] Delaying the administration of the follow-up survey beyond one year would have had the drawback of increasing attrition from the panel.

our explanations of the intriguing male-female differences in the point estimates of the labor supply response can be only suggestive.

Male and female sample members were equally likely to recall the intervention, ruling out different exposure to the treatment as an explanation. One dimension, however, along which men and women have a striking difference is the length of their work history. Only 8.7 percent of the men in our sample have a work history of less than 35 years, whereas this figure is 36.5 percent for women. The returns to working an additional year drop sharply after 35 years of earnings because at that point an additional year of earnings generally replaces an earlier year of earnings (rather than a 0) when calculating the PIA, which is based on the 35 highest years of indexed earnings. Thus, women are four times more likely than men to have a year of additional earnings count fully in the determination of their PIA. This finding implies that a better understanding of the 35-year rule would have an impact on the perceived returns to working additional years that is stronger for women than for men. Table 5 demonstrated that our intervention significantly increased knowledge of the 35-year rule, both in the entire sample as well as in the subsample of women. Thus, it is possible that the increased understanding of the 35-year rule in combination with the higher fraction of women with a work history less than 35 years caused a disproportionate increase among women in the perceived return to working longer.

If the differences in work history were responsible for the stronger treatment effects on perceived work incentives and on actual labor force participation for women, we would expect the treatment effects to be concentrated among females with a work history of fewer than 35 years. This indeed appears to be the case. The treatment effect on perceived labor supply incentives, as measured by the "better deal" question, is 0.119 (s.e. 0.064) for women with a work history less than 35 years but only 0.033 (s.e. 0.046) for those with a work history of 35 years or more. Though this difference is not statistically significant (p-value 0.203), the size and sign of the difference offer some suggestive support for the notion that the gender difference in treatment effects can traced in part to differences in work history. Similarly, we would expect the treatment effect on labor force participation to be concentrated among women with a work history of fewer than 35 years. This also appears to be the case. The treatment effect on labor force participation is 0.099 (s.e. 0.056) for women with less than 35 years of work but is 0.057 (s.e. 0.041) for those with a longer work history. While the difference in treatment effects is again not statistically significant (p-value 0.560), its sign and size are suggestive of a role of gender differences in earnings histories as part of the explanation for the stronger treatment effects on women.

However, the difference in career lengths is unlikely to be the entire story. The point estimate of labor supply response of women with 35 years of work is still much larger than that for men (though it is not significantly different from zero). Moreover, the intervention did not change the fraction of women who expected to receive benefits on their own records (83 percent). Thus, another interpretation is that women as a whole have gotten a particular message about Social Security based on the historical experience of typical women—and that this has rubbed off even on women who worked for 35 years or more. In particular, some women in our sample may have believed that as secondary earners, they get little or no marginal Social Security benefits from additional work. This would have been true for most women who retired twenty or thirty years ago, but for women today who are working into their 60s, we estimate that 70 percent or more are receiving benefits based on their own earnings record.[27] Both the information and gestalt of our intervention may have countered the historical message that Social Security's work incentives are weaker for women.

V. Conclusion

The field experiment described in this paper demonstrates that a relatively mild informational intervention can have important impacts on the labor force participation of older individuals. In particular, we find that our intervention increased labor force participation in our sample by 4 percentage points. While we found a labor supply response on average, we note that the provision of information can improve decision making even if it does not change mean outcomes. In particular, if prior perceptions of labor supply incentives were centered around the true values, better knowledge could cause an equal number of people to increase and decrease their retirement dates. Liebman and Luttmer (2012) find that for many features of Social Security, median perceptions are quite close to the true values, even though there is a wide spread around the median. Thus, it is possible that our intervention caused more than 4 percent of our sample to change labor supply, but that our estimate only picked up the net effect of increases and decreases in labor supply.

We also find that the intervention affected people's perceptions of the returns to remaining in the labor force and that both the labor supply impacts and the impacts on perceptions were concentrated in the same population subgroups—suggesting that the behavioral response may be attributable in part to updated perceptions of incentives from Social Security. We do not, however, have a conclusive explanation for why the labor supply impacts appear to be concentrated among female sample members. We speculate that the intervention may have countered perceptions that women receive little or no additional Social Security benefits from incremental labor effort, perceptions left over from an earlier era when most women received benefits based on their husbands' earnings records.

This study was designed to answer the threshold question of whether an easily scalable information intervention could alter behavior. The intervention therefore delivered information in several different mutually reinforcing ways. Before drawing policy implications from this intervention and other similar interventions in which information provision alters behavior, it will be important to learn more about the mechanisms through which such interventions produce their effects. To

[27] To generate a rough estimate of the share of women receiving benefits on their own record, we analyzed SSA's 2004 Benefit and Earnings Public-Use File. To mimic the characteristics of our sample, we identified recent retirees who had been working when they were between the age of 60 and 65 and examined whether they were receiving only retired worker benefits or whether they were also (or only) receiving benefits based on the record of a current or former spouse. Specifically, we examined female Social Security beneficiaries who were between the age of 66 and 71 in 2004 and who had positive labor earnings in 1998 (when their ages ranged from 60 to 65). In this sample, 65 percent of the women were receiving benefits based only upon their own earnings record. However, this cohort (born around 1935) was born about ten years earlier than our survey sample (born around 1945). Eckstein and Lifshitz (2011) show that female labor force participation rates rose by more than 10 percentage points at most ages between the 1935 and 1945 birth cohorts. Thus, it appears likely that the fraction in our sample who will receive benefits on their own record exceeds 70 percent. Moreover, 83 percent of the female respondents report that their benefits will be based only on their own earnings record.

the extent that informational interventions affect behavior by educating sample members and allowing them to make better choices, such interventions unambiguously raise sample member welfare. In contrast, if interventions have their impact by delivering a general message such as "working to older ages is better," then the normative implications are more complicated since the message may or may not be accurate for a particular sample member. Because our experimental mechanism allows for complete researcher control over the message delivered to sample members, it would be straightforward to conduct a follow-up study that delivered only subsets of the intervention so as to determine the relative importance of each form of information as well as the existence of interaction effects from reinforcing the message by providing information in multiple ways.

REFERENCES

Acemoglu, Daron, and Amy Finkelstein. 2008. "Input and Technology Choices in Regulated Industries: Evidence from the Health Care Sector." *Journal of Political Economy* 116 (5): 837–80.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481–95.

Benítez-Silva, Hugo A., Berna Demiralp, and Zhen Liu. 2009. "Social Security Literacy and Retirement Well-being." University of Michigan Retirement Research Center Working Paper 2009–210.

Bernheim, B. Douglas. 1988. "Social Security Benefits: An Empirical Study of Expectations and Realizations." In *Issues in Contemporary Retirement*, edited by Rita Ricardo-Campbell and Edward P. Lazear, 312–45. Stanford: Hoover Institution Press.

Bernheim, B. Douglas, and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *Quarterly Journal of Economics* 124 (1): 51–104.

Beshears, John, James J. Choi, David Laibson, and Brigitte C. Madrian. 2008. "How Are Preferences Revealed?" *Journal of Public Economics* 92 (8–9): 1787–94.

Brown, Jeffrey R., Arie Kapteyn, and Olivia S. Mitchell. Forthcoming. "Framing and Claiming: How Information-Framing Affects Expected Social Security Claiming Behavior." *Journal of Risk and Insurance*.

Chan, Sewin, and Ann Huff Stevens. 2008. "What You Don't Know Can't Help You: Pension Knowledge and Retirement Decision-Making." *Review of Economics and Statistics* 9 (2): 253–66.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99 (4): 1145–77.

Chetty, Raj, and Emmanuel Saez. 2013. "Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients." *American Economic Journal: Applied Economics* 5 (1): 1–31.

Cutler, David M., Jeffrey B. Liebman, and Mark Shepard. 2013. "An Expanded Model of Health and Retirement." Unpublished.

Delavande, Adeline, and Susann Rohwedder. 2008. "Eliciting Subjective Probabilities in Internet Surveys." *Public Opinion Quarterly* 72 (5): 866–91.

Dominitz, Jeff, Angela Hung, and Arthur van Soest. 2007. "Future Beneficiary Expectations of the Returns to Delayed Social Security Benefit Claiming and Choice Behavior." Michigan Retirement Research Center Working Paper 2007–164.

Duflo, Esther, William Gale, Jeffrey Liebman, Peter Orszag, and Emmanuel Saez. 2006. "Savings Incentives for Low- and Middle-Income Families: Evidence from a Field Experiment with H&R Block." *Quarterly Journal of Economics* 121 (4): 1311–46.

Eckstein, Zvi, and Osnat Lifshitz. 2011. "Dynamic Female Labor Supply." *Econometrica* 79 (6): 1675–1726.

Goda, Gopi Shah, Colleen Flaherty Manchester, and Aaron Sojourner. 2012. "What Will My Account Really Be Worth? An Experiment on Exponential Growth Bias and Retirement Savings." National Bureau of Economic Research (NBER) Working Paper 17927.

Gustman, Alan L., and Thomas L. Steinmeier. 2005a. "Imperfect Knowledge of Social Security and Pensions." *Industrial Relations* 44 (2): 373–97.

Gustman, Alan L., and Thomas L. Steinmeier. 2005b. "The Social Security Early Retirement Age In A Structural Model of Retirement and Wealth." *Journal of Public Economics* 89 (2–3): 441–63.

Hastings, Justine S., and Lydia Tejeda-Ashton. 2008. "Financial Literacy, Information, and Demand Elasticity: Survey and Experimental Evidence From Mexico." National Bureau of Economic Research (NBER) Working Paper 14538.

Heckman, James, Seong Hyeok Moon, Rodrigo Pinto, Peter Savelyev, and Adam Yavitz. 2010. "Analyzing Social Experiments as Implemented: A Reexamination of the Evidence from the HighScope Perry Preschool Program." *Quantitative Economics* 1 (1): 1–46.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75 (1): 83–119.

Kling, Jeffrey R., Sendhil Mullainathan, Eldar Shafir, Lee C. Vermeulen, and Marian V. Wrobel. 2012. "Comparison Friction: Experimental Evidence From Medicare Drug Plans." *Quarterly Journal of Economics* 127 (1): 199–235.

Liebman, Jeffrey B., Erzo F. P. Luttmer, and David G. Seif. 2009. "Labor Supply Responses to Marginal Social Security Benefits: Evidence from Discontinuities." *Journal of Public Economics* 93 (11–12): 1208–23.

Liebman, Jeffrey B., and Erzo F. P. Luttmer. 2011. "Would People Behave Differently If They Better Understood Social Security? Evidence from a Field Experiment." National Bureau of Economic Research (NBER) Working Paper 17287.

Liebman, Jeffrey B., and Erzo F. P. Luttmer. 2015. "Would People Behave Differently If They Better Understood Social Security? Evidence from a Field Experiment: Dataset." *American Economic Journal: Economic Policy*. http://dx.doi.org/10.1257/pol.20120081.

Liebman, Jeffrey B., and Erzo F. P. Luttmer. 2012. "The Perception of Social Security Incentives for Labor Supply and Retirement: The Median Voter Knows More Than You'd Think." In *Tax Policy and the Economy*, Vol. 26, edited by Jeffrey R. Brown, 1–42. Chicago: University of Chicago Press.

Liebman, Jeffrey B., and Richard J. Zeckhauser. 2004. "Schmeduling." http://www.hks.harvard.edu/jeffreyliebman/schmeduling.pdf.

Liebman, Jeffrey, and Richard Zeckhauser. 2008. "Simple Humans, Complex Insurance, Subtle Subsidies." In *Using Taxes to Reform Health Insurance: Pitfalls and Promises*, edited by Henry J. Aaron and Leonard E. Burman, 230–52. Washington, DC: Brookings Institution Press.

Lusardi, Annamaria, and Olivia S. Mitchell. 2007. "Financial Literacy and Retirement Preparedness. Evidence and Implications for Financial Education." *Business Economics* 42 (1): 35–44.

Lusardi, Annamaria, and Olivia S. Mitchell. 2009. "How Ordinary Consumers Make Complex Economic Decisions: Financial Literacy and Retirement Readiness." National Bureau of Economic Research (NBER) Working Paper 15350.

Mastrobuoni, Giovanni. 2011. "The Role of Information for Retirement Behavior: Evidence based on the Stepwise Introduction of the Social Security Statement." *Journal of Public Economics* 95 (7–8): 913–25.

Mitchell, Olivia S. 1988. "Worker Knowledge of Pension Provisions." *Journal of Labor Economics* 6 (1): 21–39.

O'Brien, Peter C. 1984. "Procedures for Comparing Samples with Multiple Endpoints." *Biometrics* 40 (4): 1079–87.

Rohwedder, Susann, and Kristin J. Kleinjans. 2006. "Dynamics of Individual Information about Social Security." http://business.fullerton.edu/Economics/kkleinjans/Rohwedder%20Kleinjans%20Soc%20Sec%20Exp.pdf.

Song, Changcheng. 2013. "Financial Illiteracy and Pension Contributions: A Field Experiment on Compound Interest in China." Unpublished.

Westfall, Peter H., and S. Stanley Young. 1993. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment.* New York: John Wiley & Sons, Inc.

This article has been cited by:

1. Erzo F. P. Luttmer, Andrew A. Samwick. 2018. The Welfare Cost of Perceived Policy Uncertainty: Evidence from Social Security. *American Economic Review* **108**:2, 275-307. [Abstract] [View PDF article] [PDF with links]
2. Alexander Gelber, Timothy J. Moore, Alexander Strand. 2017. The Effect of Disability Insurance Payments on Beneficiaries' Earnings. *American Economic Journal: Economic Policy* **9**:3, 229-261. [Abstract] [View PDF article] [PDF with links]
3. Saurabh Bhargava, Dayanand Manoli. 2015. Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment. *American Economic Review* **105**:11, 3489-3529. [Abstract] [View PDF article] [PDF with links]