

NBER WORKING PAPER SERIES

HEADS OR TAILS: THE IMPACT OF A COIN TOSS ON MAJOR LIFE DECISIONS
AND SUBSEQUENT HAPPINESS

Steven D. Levitt

Working Paper 22487
<http://www.nber.org/papers/w22487>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2016

I would like to thank Gary Becker, Stephen Dubner, Henry Farber, Lawrence Katz, Alan Krueger, John List, Susanne Neckermann, and Chad Syverson for valuable comments. Erin Robertson did an amazing job spearheading the project. Ellen Murphy and Mattie Toma provided outstanding research assistance. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

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 They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2016 by Steven D. Levitt. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Heads or Tails: The Impact of a Coin Toss on Major Life Decisions and Subsequent Happiness
Steven D. Levitt
NBER Working Paper No. 22487
August 2016
JEL No. D12,D81

ABSTRACT

Little is known about whether people make good choices when facing important decisions. This paper reports on a large-scale randomized field experiment in which research subjects having difficulty making a decision flipped a coin to help determine their choice. For important decisions (e.g. quitting a job or ending a relationship), those who make a change (regardless of the outcome of the coin toss) report being substantially happier two months and six months later. This correlation, however, need not reflect a causal impact. To assess causality, I use the outcome of a coin toss. Individuals who are told by the coin toss to make a change are much more likely to make a change and are happier six months later than those who were told by the coin to maintain the status quo. The results of this paper suggest that people may be excessively cautious when facing life-changing choices.

Steven D. Levitt
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER
slevitt@midway.uchicago.edu

In every life, there arise difficult decisions with potentially far-reaching consequences on lifetime utility: whether to quit a job, seek more education, end a relationship, quit smoking, start a diet, etc. Expected utility maximization is the workhorse economic model for thinking about such choices. Behavioral economics offers a host of alternative descriptive models of decision making, e.g. prospect theory, hyperbolic discounting, and the sunk cost fallacy. Yet, from an empirical perspective, economics has almost nothing to say about whether or not people are actually making good choices when it comes to their most important decisions.

There is a rich lab experimental literature exploring individual decision making under uncertainty.¹ Almost invariably, however, such studies focus on low-stakes, experimenter-generated decisions: whether to take a small gamble, how much to invest in an experimenter-defined public good, whether to accept or decline another player's division of the pie in an ultimatum game, etc.

More recently, field experiments have begun to explore decision making in real-life settings (Levitt and List, 2009; DellaVigna, 2009; Bowles et al., 2001; Gneezy and List, 2006). Just as with lab experiments, however, most field experiments investigate relatively minor decisions (e.g., what quality of baseball card to offer (List, 2002), whether to respond to a solicitation letter from a charity (Falk, 2007), and when to make mail-order catalog purchases (Anderson and Simester, 2003)). These low-stakes, isolated decisions are the types of settings in which randomization is typically available to the researcher. Bigger choices are out of reach—economists don't generally have the power to randomize people into treatments that compel them to quit their jobs or leave their spouses. Even if it were possible to, say, choose 1,000 married couples from the general population and randomly force 500 of those couples to divorce, it would not be particularly informative. Such a study would tell us about the average treatment effect of divorce. What we really care about, however, is the impact on the *marginal* decision maker. It would not be surprising if getting a divorce would have a devastating impact on the infra-marginal married person. A much more interesting question is whether divorce, *ex post*, will be the right choice for someone teetering on the edge of ending a relationship.²

Even if one found such a group of individuals who are close to indifferent between remaining married and getting divorced, an *ex post* comparison of the happiness of those who do and do not make a change still would not have an easy causal interpretation, because the people who make a change will systematically differ from those who do not on many dimensions. To convincingly answer the question, a researcher would not only need to find large numbers of these marginal individuals, but also, through some sort of randomization, influence their important life choices.

That is what I do in this study. I created a website called FreakonomicsExperiments.com. On the website, individuals who are having a difficult time making a life decision are asked to answer a series of questions concerning the decision they are struggling with. Users are presented with a wide range of questions to choose from (See Appendix A for the full set of questions offered) or invited to create their own question. One choice (e.g., "go on a diet") is assigned to heads and the other choice (in this case

¹ For surveys of this enormous literature, see Camerer (1995), Smith (1994), and Chaudhuri (2011). A notable recent contribution to decision-making under uncertainty is Gneezy, Imas, and List (2015). Fox and Tversky (1995) and Becker and Brownson (1964) are other noteworthy additions to this literature.

² To answer questions like that, previous research has typically had to rely on correlational studies (e.g. Kalmijn et al., 2009 and Pedersen and Schmidt, 2014) or natural experimental variation (e.g. Gruber and Mullainathan, 2005 and Meier and Stutzer, 2007), with the usual challenges to causal inference.

“don’t go on a diet”) is assigned to tails. The outcome of the coin toss is randomized and the user is shown the outcome of the coin toss. The coin tossers are then re-surveyed two months and six months after the initial coin toss. Additionally, prior to the randomization, coin tossers are encouraged to identify a third party (a friend or family member) to verify their outcomes. The third parties are also surveyed two months and six months after the coin toss.

While it might seem implausible that anyone would come to such a website and flip a coin, much less follow the dictate of the coin toss, the results obtained speak to the contrary. In the year of data collection, over 20,000 coins were flipped. A number of results emerge from the analysis.

First, two months into the study participants show a bias towards the status quo, in the sense that people report making a change less frequently than they predicted they would before the coin toss. Six months after the coin toss, however, this bias is gone.

Second, those who report making a change in follow-up surveys are substantially happier than those who do not make a change. This is true for virtually every question asked both two months and six months later. This correlation does not, of course, necessarily imply causality. Those who make a change differ from those who do not make a change on many dimensions.

Third, the outcome of the coin toss appears to influence the actions taken. Those who flipped heads were approximately 25 percent more likely to report making a change than those who got tails. The coin toss had a roughly equal impact on decisions across the entire range of self-stated *ex ante* likelihoods of making a change (i.e. the coin toss matters whether before the toss the coin-flipper says he/she has a 20 percent, 50 percent, or 90 percent likelihood of making the change). The coin toss was roughly equally influential on men and women, the old and the young, and across income levels. The coin toss, not surprisingly, had the biggest impact on relatively unimportant decisions like whether or not to go on a diet, but also influenced much more important choices like job quitting and ending relationships. The coin toss only influenced decisions made within the first two months of the coin toss; later changes were unrelated to the outcome of the toss.

Fourth, when it comes to “important” decisions (e.g. job quitting, separating from your husband or wife), making a change appears to be not only correlated with increased self-reported happiness, but also causally related, especially six months after the coin toss.³ Those who were instructed by the coin toss to make a change were both more likely to make the change (as noted above) and, on average, report greater happiness on the follow-up surveys. This finding is inconsistent with expected utility theory; those who are on the margin should, on average, be equally well off regardless of the decision

³ Richard Easterlin was one of the first economists to be widely recognized for work with self-reported happiness data, and since his contribution in 1974 on the link between income and subjective happiness many others have made use of such data. Dolan et al. (2008) and Frey and Stutzer (2002) provide overviews of the use of self-reported happiness data in the economics literature. Additional applications of happiness data in the field are outlined by Di Tella and MacCulloch (2006) who conclude that, treated with caution, the data have the potential to add value to empirical work. Researchers differ in their level of optimism regarding the validity of such data — Kahneman and Krueger (2006) note that the cleanest use of self-reported happiness data would “avoid effects of judgment and of memory as much as possible” but acknowledge that subject to these limitations such data can add important contributions to the field, while Bertrand and Mullainathan (2001) offer skepticism in noting that the use of a dependent variable that relies on self-reported happiness data can be problematic because “the measurement error appears to correlate with a large set of characteristics and behaviors”.

they make. Choices on “less important” decisions (e.g. dying hair, improving posture) do not generally have a measurable impact on later happiness.

All of these results are subject to the important caveats that the research subjects who chose to participate in the study are far from representative, there may be sample selection in which coin tossers complete the surveys, and responses might not be truthful. I consider a wide range of possible sources of bias and where feasible explore these biases empirically, concluding that it is likely that the first-stage estimates (i.e. the effect of the coin toss on decisions made) represent an upper bound. There is less reason to believe, however, that there are strong biases in the 2SLS estimates (i.e. the causal impact of the decision on self-reported happiness).

The structure of the remainder of the paper is as follows. Section II describes in greater detail the experiment and how it was carried out. Section III reports the results of the experiment. Section IV explores how a variety of potential biases might influence the inferences drawn from the study, and also considers how likely those biases are to be important. Because this study differs in substantial ways from standard experimental interventions by economists, the issues of bias that arise are not the typical ones economists are used to thinking about. Section V concludes.

Section II: Experimental Design

The experiment was carried out online at the website www.FreakonomicsExperiments.com.⁴ Users who arrived at the site were greeted with the home page shown in Figure 1, which offered to help people make decisions through the use of a coin flip. Those individuals who clicked “Learn More” saw the screen shot presented in Figure 2. If they proceeded further, they were shown a menu of life decisions over which to flip a coin from which they could choose; they were also given the option of designing their own customized question. After selecting a question relevant to their particular dilemma, subjects filled out a short survey that collected basic demographic data, asked them to rate their current level of happiness, probed them about the decision they were having trouble making, and gave them the opportunity to identify a third party, typically a friend or family member, who could be surveyed in the future regarding their decision.⁵ Approximately 30 percent of subjects provided the name and email address of a third party. This sub-sample of the data is of particular interest for two reasons. First, naming a third party may signal greater commitment to following the coin toss. Second, the existence of a third party provides an independent source of information to verify later participant responses, as well as a source when the subject fails to respond to follow-up surveys.

The participants were then led to a page where a simulated coin tied to a randomizing algorithm was flipped and came up either heads or tails.⁶ Subjects were reminded of what action the coin toss

⁴ For a further description of the experiment and preliminary results, written for a popular audience, see Dubner and Levitt (2014).

⁵ Users were also shown, at random, a fact relevant to the decision they were about to make. For instance, those pondering whether to quit their job were told either “The number of job openings is on the rise – up by nearly 70 percent since 2009” or “Workers who dislike their jobs report lower levels of wellbeing than the unemployed. In fact, 81 percent of the unemployed report that they are happy every day compared to only 69 percent of the unhappily employed.” There are no statistically significant differences in actions associated with having seen different facts.

⁶ Before the coin toss took place, subjects were asked how likely they were to make the change. If subjects indicated that they were very likely or very unlikely to make a change, they were taken to a page telling them that

directed them to take, and if the coin toss said to make a change, they were encouraged to make that change within the next two months. In those cases where the coin toss said don't make a change, the subjects were told to maintain the status quo for at least the next two months (e.g., if the coin toss said not to quit one's job, the subjects were asked to remain at the job for at least two months). In most, but not all cases, heads was associated with making a change and tails was associated with maintaining the status quo. For simplicity in exposition, I refer to heads in what follows as meaning that the coin toss recommended a change.

Subjects were aware that they were part of an experiment and were required to explicitly give their informed consent. Both the subjects and the third parties provided by the subjects were then surveyed 2 months and 6 months after the coin toss. Survey reminders were sent via email and included a link to an online survey site where the follow-up surveys were done. In order to encourage survey completion, those who filled out the surveys were provided with small gifts that took the form of exclusive content from Freakonomics podcasts. It should be noted, however, that I intentionally made it difficult for subjects to determine the precise objective of the study. Subjects were told that their participation would "help us gain important insights into decision making." The initial survey, prior to the coin toss, asked many questions about motivations and feelings surrounding the decision. The follow-up surveys also asked a number of questions unrelated to the actual purpose of the study.

The website FreakonomicsExperiments.com was launched on January 23, 2013. Recruiting was done through a variety of online and traditional media avenues including reddit.com, the Freakonomics podcast, the Freakonomics blog, Marginal Revolution, and articles published in The Financial Times and Forbes. Data collection at the site remained active for roughly a year, after which a scaled down version of the site remained operational, but all survey activity ended.

During the time of the study, there were approximately 165,000 unique visitors. Roughly 23,500 coin tosses took place. Excluded from the analysis are coin tosses with technical problems (primarily as a result of the user providing a faulty email address), leaving 22,511 usable coin tosses.

The distribution of these coin tosses across questions is presented in Table 1. Questions are divided into two categories corresponding to the importance of the decision for a person's life. This classification is based on a survey of individuals who were not part of the original experiment.⁷ I use this classification to aggregate questions later in the paper. "Important" questions are listed first in the table, followed by "less important" questions. Of the important questions, the single most popular was "Should I quit my job?" which attracted 2,186 coin tosses. The other "important" questions which yielded more than 1,000 coin flips were "Should I break up with my significant other?" and "Should I go back to school?"

it seemed like they had already made up their mind. Those subjects then had the option of proceeding to the coin toss or exiting.

All users were given the choice of having their outcome determined by a single coin toss, or could opt for a "best two out of three." Approximately 56 percent of users chose the "two out of three" option. In terms of subsequent behavior, there are no clear differences between those who went for the single coin versus best of three option. In what follows, I use the shorthand of a coin toss to refer to both of these options.

⁷ Respondents were asked to rate the importance of each life decision on a scale from 1-5. The correlation in rankings across individuals is quite high, with an average pairwise correlation of .707. The cutoff between "important" and "less important" is by necessity somewhat arbitrary. There was a large gap in ratings between "Should I move?" (average rating of 3.45) and "Should I go on a diet?" (average rating of 3.0), so I divided the sample there.

Among “less important” questions, over 3,000 individuals created their own questions. I mostly ignore these questions in the analysis that follows. Other popular choices related to splurging and going on a diet.

Online surveys of both the participants and the third parties were conducted two months and six months after the coin toss. The surveys of coin tossers reminded the recipient which question had led to a coin being tossed (but did not remind them of the outcome of the coin toss), and then asked, among other questions, (a) whether he/she remembered what the outcome of the coin toss had been, (b) whether an action had been taken since the coin toss, and (c) asked him/her to report their current happiness on a scale of 1-10. Third parties were asked a parallel set of questions, appropriately rephrased. For questions where a decision was essentially permanent (e.g. quitting a job), subjects were asked whether they had taken the action. On topics for which a change was potentially temporary (e.g. attempting to quit smoking which might succeed or fail), we asked subjects whether the attempt had been made.

Figure 3 reports the degree of success in obtaining follow-up surveys. There is at least one completed survey from roughly 58.34 percent of the coin tossers who did not name a third party. Those who named a third party before the coin toss were more likely (77.39 percent) to complete at least one survey, consistent with the conjecture that naming a third party signals commitment to the experiment. Adding in the surveys filled out by the third parties, I have at least one follow up survey for 83.57 percent of the coin tossers who named a third party. Response rates were higher for the two-month survey (a total of 13,935 completed surveys) than the six-month survey (8,159 completed surveys). Throughout the analysis, except where noted, I analyze the two-month and six-month samples separately.

Section III: Results

There are two questions of primary interest: (1) Did the coin toss influence behavior? and (2) What can be learned about the impact of choices on subsequent happiness? I begin with an analysis of the first question before turning to the second question. In this section, I simply report the data generated by the experiment and the treatment effects that arise from those data. There are many potential sources of bias that might arise as a result of survey non-response and untruthful responses on the part of subjects. I defer careful consideration of these potential biases to Section IV.

3.1) *Did the outcome of the coin toss influence behavior?*

Figure 4 presents data on the rate of coin toss adherence among survey respondents. The green bars correspond to two-month responses; blue represents data from the six-month survey. The values reported in the columns are the percentage of coin tossers whose actions correspond to the dictate of the coin toss, i.e making a change if heads came up and maintaining the status quo if tails was the outcome.⁸ If the coin toss has no impact on behavior, then fifty percent of the actions taken should match the coin’s dictate. The first two bars in Figure 4 reflect data from all coin tosses. After two months, roughly 63 percent of the respondents’ actions match the recommendation of the coin toss.

⁸ For those cases where I have survey responses from both the coin tosser and the third party, and they disagree as to what action was taken, I use the stated action of the coin tosser.

This implies that 13 percent of all actions were affected by the coin toss, i.e. that someone who got heads was 26 percentage points more likely to have made a change than someone who got tails. The corresponding numbers, here and in the remainder of the paper, are slightly lower at six months. This implies that some part of the impact of the coin toss is to accelerate changes that would have happened anyway, but at a later date.

The remaining bars in Figure 4 divide the sample between “important” and “less important” questions, as defined above. On “important” questions, the rates of reported coin-toss adherence are much lower than for the full sample (56.10 percent at two months; 55.82 percent at six months), but still above fifty percent. For “less important” questions, more than 67 percent of the subjects report following the coin toss at 2 months.

Prior to the coin toss, participants were asked to report how likely they believed they were *ex ante* to take the action associated with their coin toss, e.g. to propose to their significant other. They were given a menu of choices ranging from 0 to 100 percent at 10 percent intervals.⁹ Figure 5 plots the impact of the coin toss as a function of these *ex ante* likelihoods. The horizontal axis corresponds to the participants’ stated likelihood of taking an action, prior to tossing the coin. The vertical axis is the percentage of subjects who report taking the action on the two-month survey. The two lines plotted in the figure correspond to those whose coin tosses came up heads and tails respectively. A number of insights emerge from the figure. First, the outcome of the coin toss exerted influence across the entire distribution of *ex ante* probabilities. This can be seen in the fact that the line corresponding to heads is above the line for tails across the entire span of the graph by an average of roughly 20 percentage points. The coin toss had the smallest impact (i.e. the two lines are closest together) when the self-proclaimed likelihood of a change was small. A second fact that emerges from the figure is that the lines in the graph slope upward, meaning that the *ex ante* probabilities are correlated with actual actions. The predictions by the subjects are not particularly accurate, however, as the slopes of the lines are well below the 45 degree line. A non-trivial share of those who said that they would take a particular action (or non-action) with certainty did the opposite. Finally, there is some evidence of a bias towards inaction in the two-month survey data. Since roughly half the participants got heads and half tails, the overall likelihood of taking the action falls halfway between the two lines in the figure. For *ex ante* probabilities above 30 percent, the actual rate at which the action is taken is less than was predicted by the individuals. The gap is most extreme among those who predicted they would make a change with 100 percent certainty. In fact, only about 80 percent of those participants made a change in response to heads, and less than half actually changed when the coin came up tails.

Figure 6 is identical to Figure 5, except that it shows results for the six-month survey rather than the two-month survey. The general patterns observed are similar, with one notable difference. Any evidence of a bias towards inaction has disappeared. Overall, after six months, the action is taken slightly more frequently than predicted *ex ante* by the participants.¹⁰ It should be noted, however, that

⁹ The average predicted probability of taking the action across the research subjects was 41.94%. 8.38 percent predicted that there was no chance of changing; 2.58 percent thought they would change for sure. The most popular response was 50 percent.

¹⁰ Appendix Figures 2-5 mirror Figures 5 and 6, but divide the sample into “important” and “less important” questions. The same patterns are present, except that the gap between the lines for “important” questions is smaller throughout because of the reduced influence of the coin toss.

the *ex ante* probabilities refer to the likelihood of making a change within two months, not within six months.

Figure 7 shows the impact of the coin toss on actions across individual questions. Included in the figure are the results for every question with at least 150 responses. The top portion of the figure reports findings for the questions deemed “important;” the bottom part of the figure corresponds to “less important” decisions. The values reported in the figure are the percentage of all respondents to the two-month survey who report taking the action that corresponds to the coin outcome. With the exception of “Should I move?” which shows no impact of the coin toss, for all the other “important” choices between 55 and 60 percent of the subjects report following the suggestion of the coin on the two-month survey. Decisions on “less important” questions, as might be expected, are more affected by the coin toss, with the highest compliance rate on “Should I break my bad habit” (over 80 percent), “Should I go on a diet,” “Should I quit drinking,” and “Should I try online dating.” Appendix Figure 1 is identical to Figure 7, except that it shows results for the six-month survey rather than the two-month survey. The patterns are similar.

All of the numbers presented thus far are raw data. Table 2 demonstrates that the impact of the coin toss is both robust to the inclusion of covariates and is highly statistically significant. Each column of Table 2 reports the results of a linear probability model in which the dependent variable is a dichotomous variable corresponding to whether the survey respondent says a change was made. Included as right-hand-side variables are the result of the coin toss, how likely the subject said they were to change *ex ante*, a range of demographic variables, whether the subject opted for the “best two out of three coin toss” option, and an indicator variable for the particular question for which the coin was tossed. Columns 1 and 2 reflect the whole sample. Columns 3 and 4 are the subset of “important” questions, and columns 5 and 6 correspond to the “less important” questions. The top row is the coefficient on the coin toss coming up heads. For all questions on the two-month survey, individuals who got heads report being 24.9 percentage points more likely to have made a change than those who got tails. This result is highly statistically significant. The point estimate at six months is slightly smaller (.211), implying that some of the impact of getting heads operates through accelerating the timing of a change. Comparing important questions (columns 3 and 4) to less important questions (columns 5 and 6), the impact of the coin toss is only about one-third as large for important questions, but is still highly statistically significant. The coin-tosser’s *ex ante* assessment of how likely he or she is to make a change is also highly informative about whether a change is eventually made. If the subjects made unbiased forecasts, the coefficient on this variable would be one; in actuality it ranges between .279 and .597. Subjects are better predictors of their own behavior on important questions than on less important ones. The only other variable which has a strong and consistent relationship to making a change is age. Older subjects are less likely to make changes, especially on important questions.

3.2) Is there a causal impact of making a change on happiness?

The results above suggest that the outcome of the coin toss affected the behavior of some participants. Consequently, the coin toss has the potential to shed light on the question of whether making a particular change (e.g. going on a diet) has a positive or negative impact on self-reported happiness. Before the coin toss, those who will get heads are, in expectation, identical in all respects to those who will get tails. If the only channel through which the coin toss operates is to increase the likelihood that the particular change in question is made, then the coin toss can serve as an instrumental variable.

More formally, let H represent happiness which is influenced by the choice of whether or not to take some binary action A . Additionally, let the set of all other factors that influence H be captured by some vector of variables X . For instance, relevant X 's might include the salary of one's current job, what city one lives in, the level of education, how happily married one is, etc. Some of these X 's might be observable, but many would not be. A simple comparison of happiness amongst those who take the action ($A=1$) versus those who don't ($A=0$), i.e.

$$E[H|A=1] - E[H|A=0]$$

is unlikely to have a causal interpretation because X is not held constant across those who do and don't switch jobs. Empirically, those who make a change are statistically significantly younger, less likely to be married, less educated, and lower income than those who do not make a change. While it is possible to control for these observable factors, it is likely that these two groups differ substantially on unobservable dimensions as well. *A priori*, the sign of the bias in OLS is not obvious.

OLS suffers from a second weakness: a simple comparison of *everyone* who quits their job to *everyone* who doesn't quit their job doesn't answer the economically interesting question. When considering the impact of making a change, it is the *marginal* actor who is of primary interest. There are many happily married couples and a few that are so disastrously unhappy that divorce is certain. A comparison of these two sets of couples tells us nothing about how getting divorced will affect the happiness of the couples who are truly marginal.

The outcome of the coin toss, used as an instrumental variable, potentially solves both of those problems. Let C represent an indicator variable corresponding to one if the coin comes up heads and 0 otherwise. Under the assumptions that

$$E[A|C=1] - E[A|C=0] \neq 0 \text{ and}$$

$$E[X|C=1] - E[X|C=0] = 0$$

then a simple Wald estimator provides an estimate of the causal impact of action A on happiness H :

$$\hat{B}_{Wald} = \frac{E[H|C=1] - E[H|C=0]}{E[A|C=1] - E[A|C=0]}$$

As long as the only channel through which the coin toss operates is via influencing the likelihood that the action in question is taken, then the Wald estimator represents a local average treatment effect on H of taking the action A , for that group whose behavior is influenced by the coin toss, i.e. the people who are so marginal that they are willing to have their action swayed by a coin toss.

Table 3 presents the basic empirical findings regarding the link between choices and self-reported happiness. Columns 1-3 correspond to the two-month survey; columns 4-6 reflect the six-month survey. Columns 1 and 4 are first-stage estimates, i.e. how much did the coin toss affect the action taken. Columns 2 and 5 are OLS estimates, which show the extent to which those who make a change are more or less happy than those who maintain the status quo. The OLS estimates are explicitly correlational – to the extent that people who do and do not make a change differ systematically, the OLS estimates will not have a causal interpretation. Columns 3 and 6 are the instrumental variable estimates. Under the assumption that the only channel through which the outcome of the coin toss affects happiness is through the choice made, the instrumental variable estimates capture the causal impact of the action on

self-reported happiness. The first row of the table presents results aggregated across all the questions. The second and third rows also report aggregated data, but classifying questions as either “important” or “less important.” The remaining rows of the table report results for individual questions.¹¹ Each entry in the table is from a different regression. Only the key coefficient of interest is presented in the table. In all specifications, I include a basic set of control variables mirroring those included in the first-stage regressions reported earlier. Full results are available in an online appendix.

The OLS results, presented in columns 2 and 5, carry a consistently positive sign at both two months and six months. 39 of the 46 OLS coefficients are positive, many of them statistically significantly so. This means that those who make a change report increased happiness relative to those who maintain the status quo. For important questions, the difference in reported happiness is especially large – around one full point on a 10 point scale, or nearly two-fifths of a standard deviations. As argued above, however, these OLS coefficients need not imply causality.

Indeed, the instrumental variable estimates tell a much more nuanced story than do the OLS estimates. Focusing first on the data that aggregate across all questions, at two months, the 2SLS coefficient is very close to zero and not statistically significant. This stands in contrast to the OLS results. I can reject the null hypothesis of equality of the OLS and 2SLS coefficient. At six months, however, both OLS and instrumental variables tell a similar story: change is good. Those individuals whose coin came up heads report being happier six months later than those whose coin came up tails. The instrumental variable coefficient in the top row of column 6 implies a causal impact of increased happiness of .476 on a 1-10 scale, or over one-fifth of a standard deviation of the within person change in happiness over six months.

Rows two and three of the table show that all of the impact on happiness in the 2SLS estimates comes from making a change on the “important” questions. At two months, “important” questions carry a positive, but not statistically significant coefficient. By six months, the 2SLS coefficient for “important” questions is large and statistically significant, implying a causal impact of one standard deviation change in reported happiness in response to making a change. For “less important” questions, the impact of a change at both time windows is small, negative, and statistically insignificant despite being relatively precisely estimated.

The remaining rows of Table 3 present results for individual questions. These coefficients are not precisely estimated and are statistically significant in only a few instances. Job quitting and breaking up both carry very large, positive, and statistically significant coefficients at six months. Going on a diet is positive and statistically significant at two months, but has a small and insignificant impact by six months. Online dating is positive and significant at the .10 level at two months, but turns negative by six months. Splurging is negative and significant at the .10 level at two months, but has no discernible impact by six months. Attempting to break a bad habit is negative with a t-stat of 1.5 at both points in time, perhaps because breaking bad habits is so hard. For those subjects who reported trying to break a

¹¹ I limit the sample of questions to those in which the coin tossers are making a choice between a change and the status quo. This eliminates questions like “Should I attend college A or college B?” Since colleges A and B are different across people, it is difficult to know how to evaluate such questions. The same is true with the widely varying “create your own” questions, which are also excluded.

bad habit, third parties said the bad habit had actually been broken only 20.93 percent of the time at two months and only 24.49 percent at six months.

Table 4 explores the sensitivity of the aggregated “important” question estimates across subsamples of the data.¹² The columns in Table 4 match those of Table 3. The top row of Table 4 replicates the second row of Table 3, which serves as the baseline specification against which the other results of Table 4 can be compared. Relatively few strong patterns emerge in Table 4. With respect to the first stage, the only large difference that emerges is that (as expected) those who report being likely to follow the coin toss are, indeed, about three times more likely to follow the coin toss. For the OLS estimates, older subjects have a greater increase in reported happiness from changes than younger subjects, as do people whose baseline happiness is low. On the age dimension, this pattern is interesting because older subjects are less likely to make changes than younger ones. There are few discernible patterns in the 2SLS comparisons, in large part because of imprecision. The one breakdown that shows a large differential in point estimates that persists across periods is a greater benefit to change for high-income subjects relative to low-income subjects.

Section IV: Potential biases

There are many potential biases in the results presented above. The sources of bias fall into three broad categories: non-representativeness of the subject pool, selective response to the surveys, and untruthful answers to the survey questions. I tackle these three sets of concerns in turn, in each instance considering how the biases might affect both the first-stage estimates (i.e. the willingness to follow the coin toss) and the instrumental variable estimate of the causal impact of taking an action on future self-reported happiness. It is important to note that many stories that might at first blush seem likely to bias the results (e.g. happy respondents are more likely to complete surveys, people who change are more likely to respond) in fact do not have a first order impact on any of the estimates. I limit the discussion below to sources of bias which, if present, will have a first order impact on the estimates.

4.1) Non-representative subject pool

There can be no doubt that the subject pool participating in this study is highly unusual. The great majority of the recruitment for the study was done through social media associated with Freakonomics, so participants are likely to both be aware of my prior research and favorably inclined towards it. Participants tended to be young, male, highly and educated. Secondly, in the recruiting for the study, I emphasized that I was only interested in people who were having a difficult time making a life decision. This was true both in the marketing to get subjects to the website, and in the messaging once subjects arrived at the site. Consequently, individuals who are on the margin are highly overrepresented, intentionally, in the subject pool. Finally, this is a group which is apparently attracted to the idea of using a coin toss to potentially resolve major life dilemmas. It is unclear whether that is a trait that is widespread in the population.

¹² Appendix Table 1 presents parallel sensitivity analysis for the “less important” questions.

All of these factors suggest that subjects in this sample are far more likely to have been influenced by the coin toss than would a randomly drawn sample, i.e. the first stage is much stronger in this group than would be the case more generally.

It is less clear, however, precisely how or why this sample selection would bias the paper's estimates of causal effects of decisions. One possible channel would be that the people who participated in this study are particularly bad at making decisions on their own. So, for instance, they might tend to have difficulty making changes and wait far too long to make changes when it is obvious that a change needs to be made, and thus accrue large improvements to happiness once change occurs. However, if that were true I would have expected to see strong positive casual effects on happiness of making a change in the two-month survey, but that does not happen.

4.2) Selective survey responses

The results presented throughout this paper are based on the subset of study participants who completed surveys. If survey respondents are not a random sample of the coin tossers, a number of different biases may be introduced, depending on the nature of the selection. The presence of the third parties identified by the subjects potentially allow me to assess both the size and direction of these possible biases.

Selective survey response can potentially affect each of the estimates presented in this paper: whether people follow the coin toss, the OLS estimates of changes on happiness, and the 2SLS estimates that use the coin toss as an instrument. I deal with these three cases in turn. Throughout this section I focus on the subset of important questions since those are the questions that yielded interesting results above.

Selective response biasing the first stage: Are those who follow the coin toss more likely to report?

The measured impact of the coin toss on making a change will be exaggerated if those who follow the coin toss are more likely to respond to the survey than those who go against it. Given that the website made it clear to participants that following the coin toss was important to me, it seems plausible that those who followed the coin toss would be more likely to respond. Those who make a change might tend to fill out the survey more often if they get heads, and those who do not make a change might complete the survey with a higher probability if they get tails.¹³

¹³ The effect of this type of selection on estimates of the causal link between making a change and subsequent happiness is more subtle. As long as the coin toss has some real impact on behavior, then the 2SLS estimates will be a mixture of that causal, randomization-induced variation and variation induced by the sample selection. If, for instance, the extra individuals who are induced to respond are (as good as) randomly drawn from the underlying subject distribution, then the 2SLS will be a mix of the true causal impact and the OLS estimate of the correlation between change and future happiness. But it is also possible that the kind of people who are very sensitive to pleasing or disappointing the experimenter are different on average than the other subjects. These subjects might feel guilty after making a change, and be worse off after the change than other participants, leading the 2SLS estimate to be too small. One could tell equally compelling stories as to how the bias could go the other direction as well.

To measure the actual degree of sample selection on this dimension requires some group of research subjects for whom I know the action they took, even if they do not complete the survey. The third parties are critical in this dimension. Conditional on a third party having completed a questionnaire, I am able to compare the likelihood the subject completes a survey as a function of whether or not they followed the coin toss (using as a proxy the third party's assessment of whether the coin toss was followed). Table 5 does precisely this. Entries in the first two columns of the table are the percentage of subjects who complete a survey, conditional on the third party's opinion as to whether the subject followed the coin toss (column 1) or did not follow the toss (column 2). The third column is the difference between the first two columns. Note that I did not actually ask the third parties whether the coin toss was followed, but rather, what action the subject took, which I then compare to the recommendation of the coin. The rows of the table correspond to the two-month and six-month surveys respectively. Starting in the upper left corner, when the third party completes a two-month survey and says the action taken matches the coin toss, approximately 86 percent of the subjects also complete the survey. The second entry in the top row shows that when the third party says the subject did not follow the coin toss, reporting rates are roughly 80 percent, or 6.1 percentage points lower as shown in column 3. All the reporting rates are lower at the six-month survey, but the relative patterns are similar, with those who followed the coin toss more likely to report. Thus, there does appear to be biased reporting along this dimension.

Assuming the same degree of sample selection observed among this set of subjects holds across the whole population, back-of-the envelope calculations suggest that about one-fifth of the estimated first-stage impact might be due to this bias.

Selective response biasing OLS: Are happy changers especially likely to report?

It is possible that those who make a change feel particular pride if things turn out well and greater shame if the change feels like a mistake *ex post*. If that is the case, then the OLS estimates of the benefit of a change are likely to be exaggerated.¹⁴

Table 6 explores this possible bias. The top panel of the table corresponds to the two-month survey; the bottom panel reflects the six-month survey. In both cases, the sample is restricted to those subjects for whom a third party survey is completed. I divide the sample of subjects according to whether the third party says the subject is above or below the average level of happiness at the time of the follow-up survey. The columns of the table reflect whether the third party believes that the subject made a change. The entries in the table are the percent of subjects in that category who complete a survey. The parameter of interest is the difference-in-difference: are changers disproportionately likely to report when happy relative to non-changers. Focusing first on the top row of the top panel of the table, among subjects judged by their third party to be above average on happiness, reporting rates are six percentage points higher (89.7 percent versus 83.7 percent) when a change is made than when no change occurred. For subjects who are below average on happiness in the eyes of the third party, the

¹⁴ Although it might seem like this type of selection would be very damaging to the interpretation of the 2SLS results as well, in actuality, it is not likely to affect things much. It has no obvious impact on the first-stage estimates, because the selection is operating on the happiness dimension, not on whether a subject made a change or not. And because this type of selection affects both those who got heads and those who got tails, the overall level of reported happiness for those who flipped heads and tails – which determines the numerator of 2SLS – is not obviously biased.

gap in reporting rates is only 1.2 percentage points. This suggests that, indeed, there is some bias at two months towards “happy changers” reporting. The same pattern, even stronger, appears in the bottom panel of the table which reflects the six-month survey.

Back-of-the-envelope calculations imply that these differences in reporting will exaggerate to the OLS estimates of making a change by roughly 10 percent on the two-month survey and roughly 20 percent on the six-month survey.

Selective response biasing 2SLS: Are happy heads and sad tails especially likely to report?

It is not obvious why people who get heads would be disproportionately likely to report if happy, but if they do, it will greatly bias the 2SLS estimates. Consequently, I explore whether this bias is present in the data in Table 7.¹⁵ This table has the same structure as Table 6. The only difference is that the columns of this table correspond to whether the subject got heads or tails. Once again, the difference-in-difference is the parameter of interest: if this bias is present, then happy heads should disproportionately report.

The numbers in Table 7 show no evidence of this form of bias. On both the two-month and six-month surveys, happy subjects are more likely to respond, but in neither case is there a notable difference between those who got heads versus those who got tails.

4.3) Untruthful answers from the subjects

In the cases considered above, sample selection is induced by differences in survey response rates across participants, but the maintained assumption is that the research subjects truthfully answer the questions that are asked. If respondents lie, this will also affect the estimates. In what follows, I consider three different types of lies that subjects might tell in follow-up surveys: (1) claiming to have followed the coin toss when that is not true, (2) exaggerating the degree of happiness after making a change, and (3) exaggerating how happy they are if they follow the coin toss. I address these three concerns in turn.

In testing for untruthful answers on the part of subjects, my approach is always the same: I compare the answers participants give to those of the third parties, under the assumption that the third parties have no reason to lie, unlike the subjects, who may be embarrassed about their actions or the consequences of their actions. Not all disagreements imply lying—third parties might not be fully informed—but to the extent that there are systematic patterns to the disagreements, this may be a sign of lying.

Do subjects claim to have followed the toss when they have not actually done so?

Subjects may feel pressure to say that they have followed the coin toss, especially because I so heavily emphasized the importance of doing so in advance of the coin being tossed.¹⁶ An obvious impact of lying of this sort is that it will exaggerate the first-stage estimates. The most likely consequence for the

¹⁵ One possible story is experimenter demand effects. If sophisticated subjects managed to (correctly) infer that the purpose of this study was to use the coin toss as a randomizing device to estimate a causal impact of making a change on future happiness, and additionally guessed (incorrectly) that I was hoping to find that change is beneficial, then in order to please me, they might have differentially reported along these dimensions.

¹⁶ Note, however, that both the two-month and six-month surveys emphasized that I only cared about the truth.

2SLS estimates will be to *understate* the true causal impact of a change. This is because the 2SLS estimate is the ratio of the difference in happiness of those flipping heads versus tails over the difference in the probability of making a change across heads and tails. The numerator is unaffected by this type of lying, but the denominator is exaggeratedly large, shrinking the 2SLS point estimate. OLS estimates of the value of making a change will also be biased towards zero because of attenuation bias associated with agents being misclassified.

Table 8 reports, for the set of subjects for whom I have survey responses from both the participant and the third party, the rate of coin toss following. Starting in the upper left corner of the table on the two-month survey, subjects report following the coin toss 61.4 percent of the time compared to 58.0 percent for third parties. The gap is smaller and reverses sign at six months. The data suggest some possibility that the two-month first stage may be exaggerated slightly (with the 2SLS estimates and OLS consequently understated), but do not support such a story for the six-month survey.

Do subjects exaggerate how happy they are when they make a change?

Although subjects do not have any particular reason to lie to the experimenter regarding how happy they are after a change, it is possible that they lie to themselves for psychological reasons. For instance, if making a change is costly (e.g. breaking up with a girlfriend), then it may be difficult for a person *ex post* to accept that the choice turned out poorly. A person may engage in self-deception not to have to feel the regret associated with the action. This sort of deception will have a first-order impact of exaggerating the OLS estimates of the impact of making a change. It will have no impact at all on the first-stage estimates, but will somewhat inflate the 2SLS estimates since a greater share of those who flipped heads will have made a change and exaggerated how happy they are.

To test for this source of bias, I estimate the basic OLS specifications of the table, but using the *third party* estimate of how happy the subject is as the dependent variable, rather than the subject's own report. The assumption underlying this approach is that third parties have no obvious reason to distort their responses.¹⁷

I report the results of this exercise in Table 9. For purposes of comparison, the first two rows of the table report results using the subject's own happiness report. The first row replicates the basic specifications reported in Table 3 for important questions. The second row is identical to the first row, except that it limits the sample to those subjects for whom there is also a third party survey. This second row is relevant because that same sample restriction is present in the third row, which uses third party assessments of happiness as the dependent variable. A comparison between the three rows shows that restricting the sample somewhat increases the measured impacts (i.e. making a change is associated with a greater increase in happiness in the subset of the population where both the subject and the third party respond), but that the results are not sensitive to whether I use the subject's own happiness as the outcome or the third party's assessment. Consequently, there is little evidence that this bias is present empirically.¹⁸

¹⁷ It is possible that the coin tosser misrepresents his or her happiness not just to the experimenter, but also to friends and family, in which case their assessment might also be biased. If that is the case, then using third party evaluations may not fully address the bias due to misrepresentation.

¹⁸ In principle, I can carry out the same exercise using the third party happiness reports as the dependent variable in the 2SLS estimates to test whether misreporting of happiness might bias the 2SLS estimates. In practice,

4.4) Summary of potential biases

Summarizing the discussion above, it is likely that the first-stage estimates in this paper are exaggerated, both because of the selected sample participating in this study and reporting biases. There is also evidence that differential reporting may bias upward the OLS estimates of making a change on subsequent happiness by 10-20 percent. There is no obvious evidence for strong bias in the 2SLS, nor does it seem to be the case that lying (as opposed to differential reporting rates) is biasing the various estimates.

Section V: Conclusion

Empirical economists are increasingly moving from a role of consumers of data to producers of data. This paper represents an extreme expression of that trend. It is difficult to imagine how one could hope to answer the questions addressed in this paper without generating the data. As the prominence of social media grows, opportunities to recruit subject pools for randomized field experiments from broad swaths of the population will only increase.

The results of this paper suggest the presence of a substantial bias against making changes when it comes to important life decisions, as evidenced by that fact that those who do make a change report being no worse off after two months and much better off six months later. The results of this paper are, of course, merely suggestive. If the results are correct, then admonitions such as “winners never quit and quitters never win,” while well-meaning, may actually be extremely poor advice.

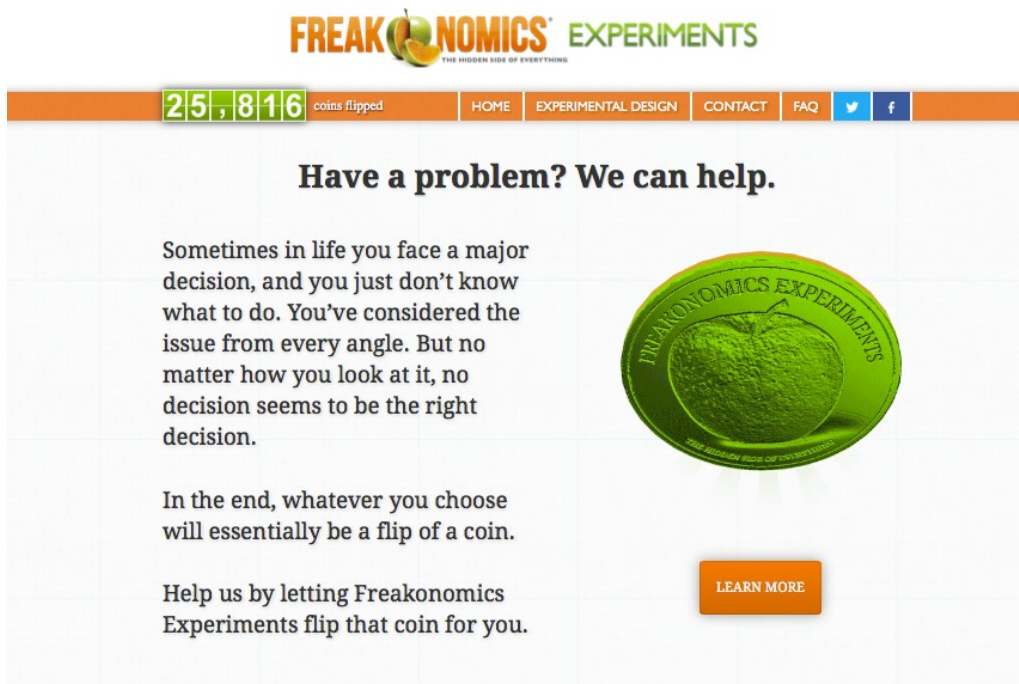
however, the estimates are so imprecise that they are uninformative. The 2SLS standard errors when I restrict the sample to cases where both the subject and the third party report are roughly one in the two month survey and nearly three in the three month survey. Thus, no reasonable hypothesis can be rejected by the data.

References

- Anderson, E. and D. Simester (2003). "Effects of \$9 Price Endings on Retail Sales: Evidence from Field Experiments," *Quantitative Marketing and Economics*, 1, 93-110.
- Becker, S. and O. Brownson (1964). "What Price Ambiguity? Or the Role of Ambiguity in Decision-Making," *Journal of Political Economy*, 72(1), 62-73.
- Bertrand, M. and Mullainathan, S. (2001). "Do People Mean What They Say? Implications for Subjective Survey Data," *American Economic Review*, 91(2): 67-72.
- Bowles, S., R. Boyd, C. Camerer, E. Fehr, E. Gintis, J. Henrich, R. McElreath (2001). "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies," *American Economic Review*, 91(2), 73-78.
- Camerer, C. (1995). "Individual Decision Making," in J. Kagel and A. Roth (eds.), *The Handbook of Experimental Economics*, Princeton University Press.
- Chaudhuri, A (2011). "Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature," *Experimental Economics*, 14, 47-83.
- DellaVigna, S. (2009). "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47(2), 315-372.
- Di Tella, R. and R. MacCulloch (2006). "Some Uses of Happiness Data in Economics," *Journal of Economic Perspectives*, 20(1), 25-46.
- Dolan, P, T. Peasgood, M. White (2008). "Do We Really Know What Makes Us Happy? A Review of the Economic Literature on the Factors Associated with Subjective Well-being," *Journal of Economic Psychology*, 29, 94-122.
- Easterlin, R. A. (1974). "Does Economic Growth Improve the Human Lot? Some Empirical Evidence," in P. David and M. Reder (eds.), *Nations and Households in Economic Growth*, New York and London: Academic Press.
- Falk, A. (2007). "Gift Exchange in the Field," *Econometrica*, 75(5), 1501-1511.
- Fox, C. and A. Tversky (1995). "Ambiguity Aversion and Comparative Ignorance," *The Quarterly Journal of Economics*, 110(3), 585-603.
- Frey, B. and A. Stutzer (2002). "The Economics of Happiness," *World Economics*, 3(1), 1-17.
- Gneezy, U., I. Imas, and J. List (2015). "Estimating Individual Ambiguity Aversion: A Simple Approach," NBER Working Paper No. 20982.
- Gneezy, U. and J. List (2006). "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments," *Econometrica*, 74(5), 1365-1384.
- Gruber, J. and S. Mullainathan (2005). "Do Cigarette Taxes Make Smokers Happier," *Advances in Economic Analysis and Policy*, 5(1), Article 4.

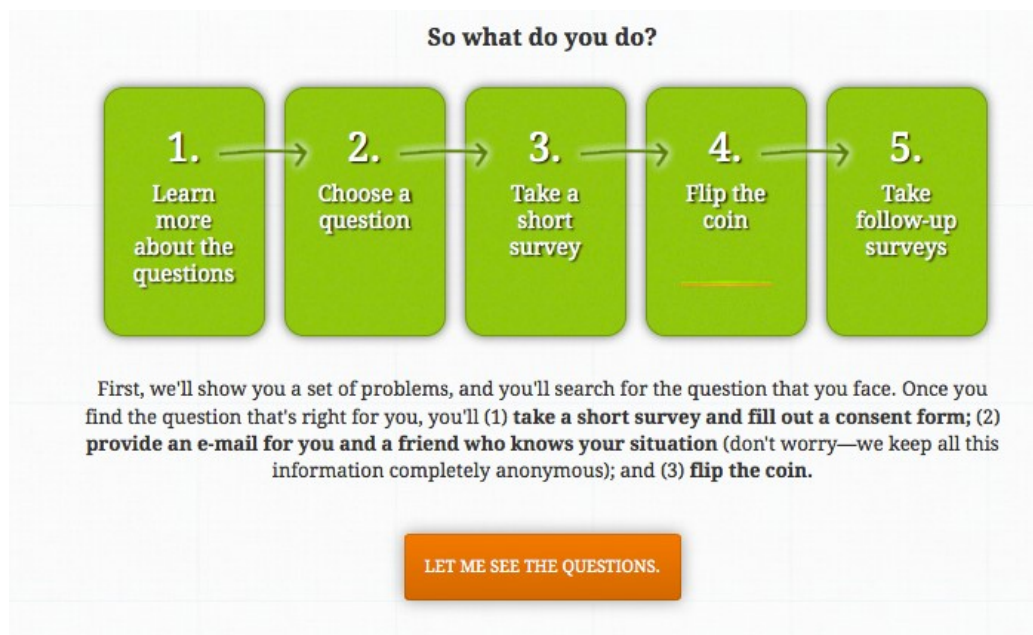
- Kahneman, D. and A. Krueger (2006), "Developments in the Measurement of Subjective Well-being," *Journal of Economic Perspectives*, 20(1), 3-24.
- Kalmijn, M., A. Liefbroer, J. Soons (2009). "The Long-Term Consequences of Relationship Formation for Subjective Well-Being," *Journal of Marriage and Family*, 71(5), 1254-1270.
- Levitt, S. and S. Dubner (2014). Think Like a Freak. New York: William Morris.
- Levitt, S. and J. List (2009). "Field Experiments in Economics: The Past, the Present, and the Future," *European Economic Review*, 53, 1-18.
- List, J. (2002). "Preference Reversals of a Different Kind: The "More Is Less" Phenomenon," *American Economic Review*, 92(5), 1636-1643.
- Meier, S. and A. Stutzer (2007). "Is Volunteering Rewarding in Itself?," *Economica*, 75(297), 39-59.
- Pedersen, P. and Schmidt, T (2014). "Life Events and Subjective Well-Being: The Case of Having Children," IZA Discussion Paper No. 8207.
- Smith, V. L. (1994). "Economics in the Laboratory," *Journal of Economic Perspectives*, 8(1), 113-131.

Figure 1. Website Home Page



Note: Figure 1 presents the Freakonomics Experiments website home page. This is the first page users saw when visiting the site.

Figure 2. Experiment Flow Chart



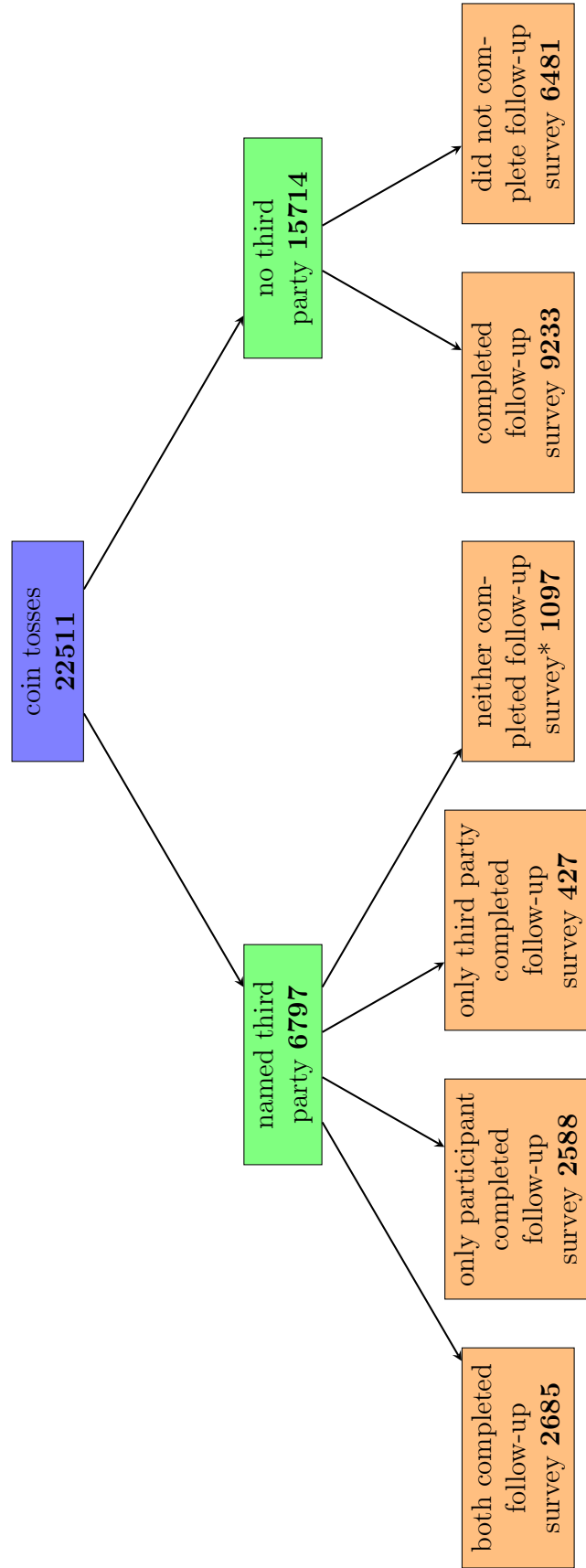
Note: Figure 2 presents a flow chart that maps out the steps that users follow when participating in the experiment. Users saw this screen after clicking the 'Learn More' button displayed in Figure 1.

Table 1. Question Attributes

Question	Number of Tosses	Important Question?	Choice Between Action and Status Quo?
Should I quit my job	2186	Yes	Yes
Should I break up	1686	Yes	Yes
Should I go back to school	1203	Yes	Yes
Should I start my own business	893	Yes	Yes
Should I move	762	Yes	Yes
Should I quit smoking	499	Yes	Yes
Should I have a child	415	Yes	Yes
Should I propose	220	Yes	Yes
Should I retire	120	Yes	Yes
Should I adopt	42	Yes	Yes
Create your own question	3485	No	No
Should I splurge	1491	No	Yes
Should I go on a diet	1134	No	Yes
Should I break my bad habit	984	No	Yes
What should I major in	959	No	No
Should I get a tattoo	876	No	Yes
Should I try online dating	699	No	Yes
What college should I go to	656	No	No
Should I join a gym	630	No	Yes
Should I dye my hair	514	No	Yes
Should I sign up for a running event	431	No	Yes
Where should I move to	425	No	No
Should I grow facial hair	424	No	Yes
Should I quit drinking	401	No	Yes
Should I ask for a raise	385	No	Yes
Should I start volunteering	364	No	Yes
Should I rent or buy	295	No	No
What school should I send my child to	130	No	No
Should I get a roommate	106	No	Yes
Which house should I buy	96	No	No

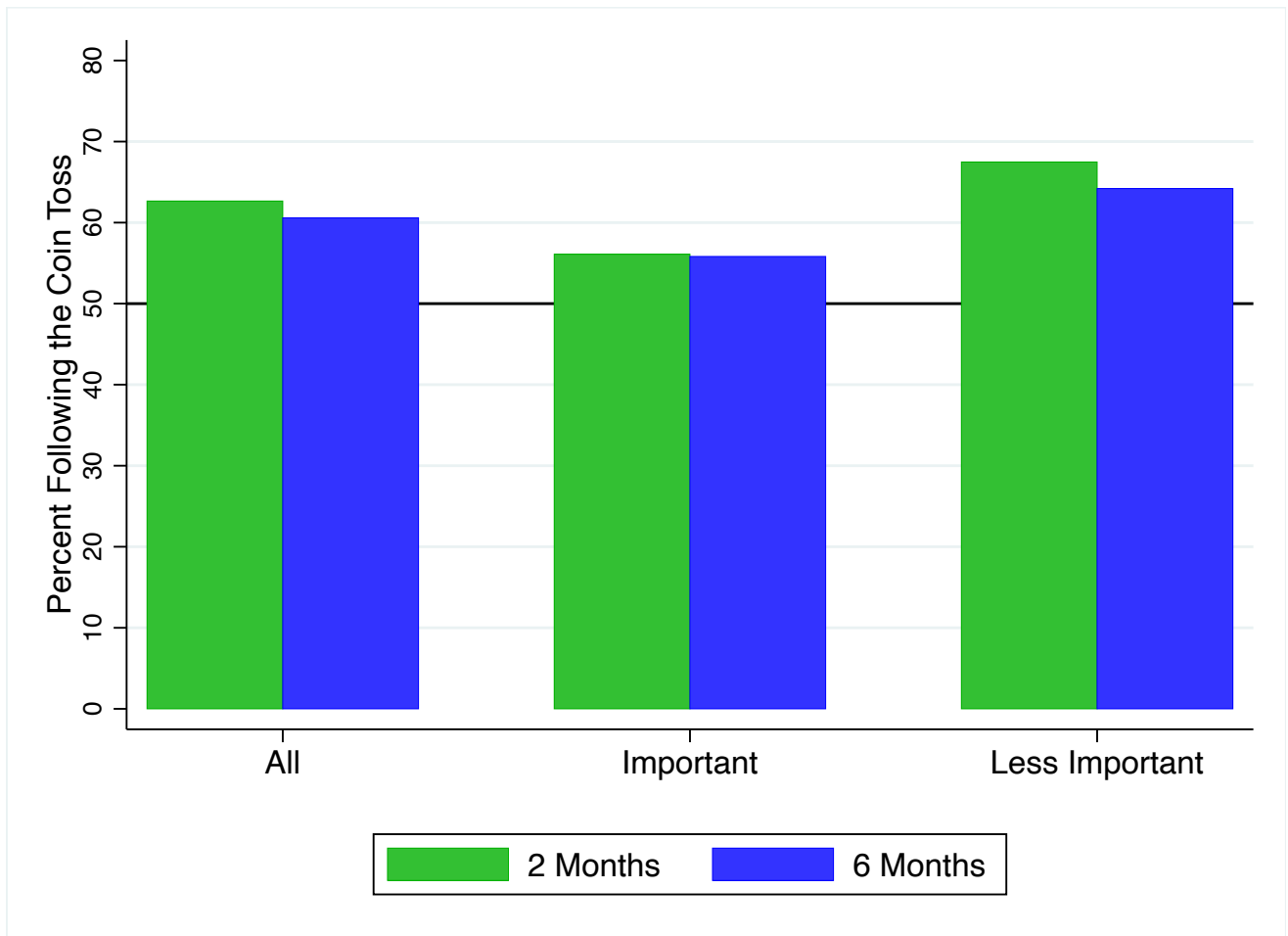
Note: This table presents summary information by question. The first column displays the number of coins tossed for each question. The second column indicates whether the question is considered an important question, where important questions are displayed in the top panel of the table. The third column indicates whether a question represents a choice between action or maintaining one's status quo (Yes) as opposed to a choice between two possible actions (No).

Figure 3. Follow-up Survey Response Rates



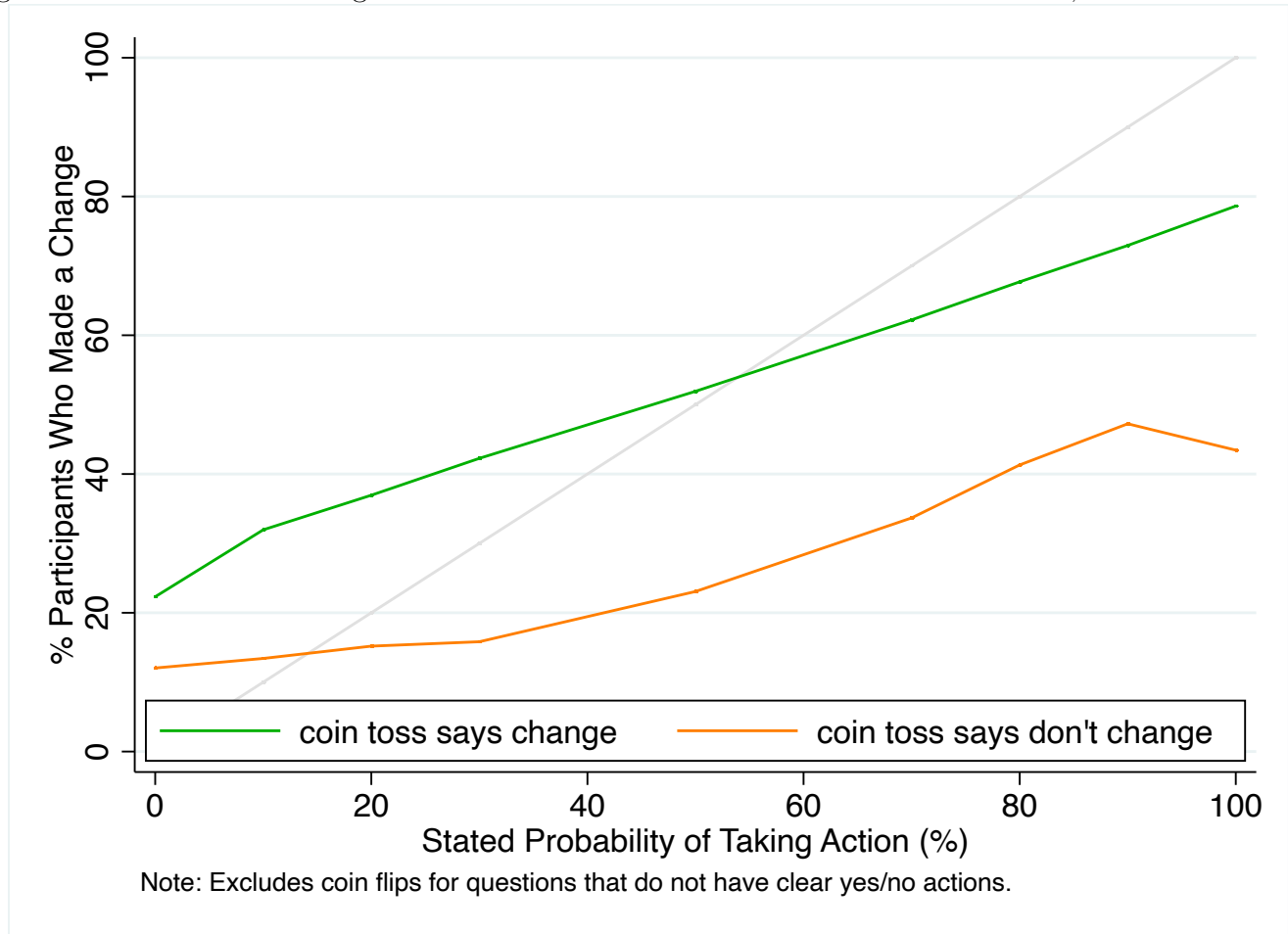
Note: Figure 3 presents the number of total tosses and the number of completed surveys according to whether a third party was named. Note that the category consisting of participants who did not complete a follow-up survey includes those who did not complete their first follow-up survey and who never received a second follow-up survey because the experiment ended before the participant would receive this second survey. 3,655 participants never received their second follow-up survey. 798 participants did not receive their first follow-up survey due to the experiments end date and were thus excluded from our analysis.

Figure 4. Coin Toss Adherence among Survey Respondents



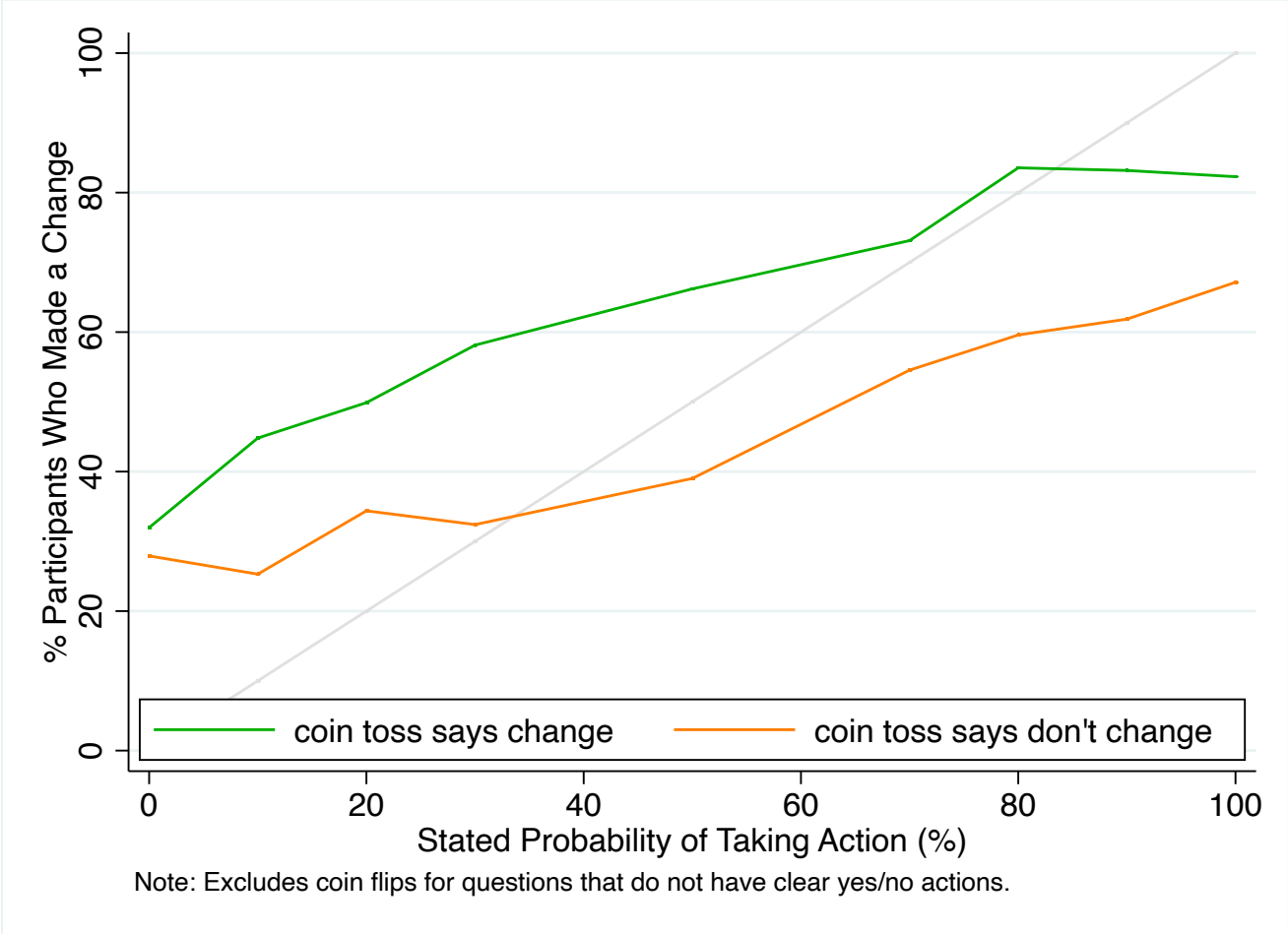
Note: Figure 4 presents coin toss adherence based on two- and six-month survey responses. The vertical axis reflects the percent following the coin toss. The horizontal axis categorizes response rates by question type and survey.

Figure 5. Likelihood of Taking Action as a Function of Ex Ante Stated Probabilities, Two-Month Survey



Note: Figure 5 presents the percent of participants who make a change by the two-month survey mark according to their stated probability of changing and the result of the coin flip. The vertical axis reflects the percent of respondents who reported making a change. The horizontal axis groups respondents according to their stated ex-ante likelihoods of making a change. Responses are categorized according to whether the coin came up heads (make a change) or tails (no change).

Figure 6. Likelihood of Taking Action as a Function of Ex Ante Stated Probabilities, Six-Month Survey



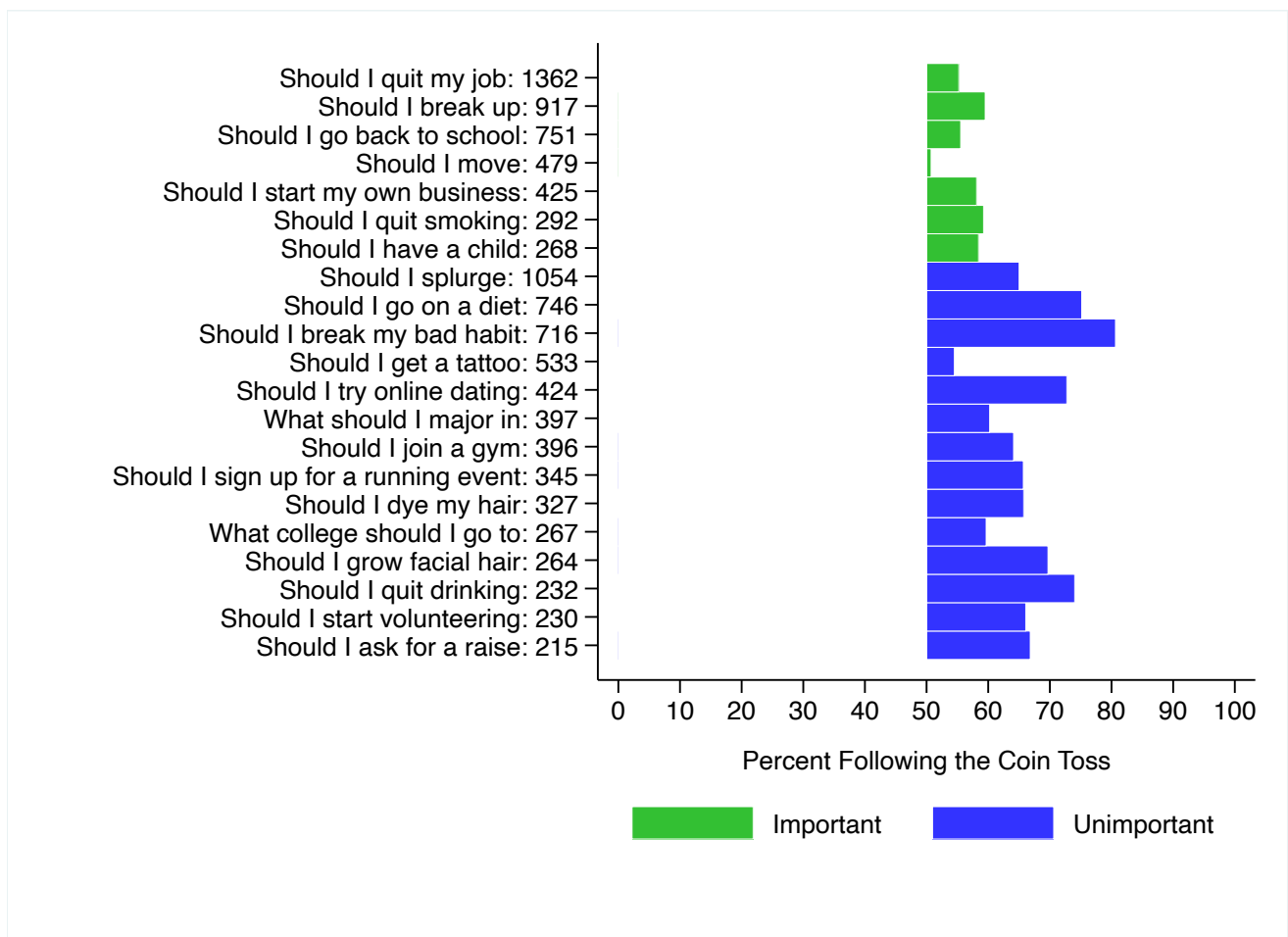
Note: Figure 6 presents the percent of participants who make a change by the six-month survey mark according to their stated probability of changing and the result of the coin flip. The vertical axis reflects the percent of respondents who reported making a change. The horizontal axis groups respondents according to to their stated ex-ante likelihoods of making a change. Responses are categorized according to whether the coin came up heads (make a change) or tails (no change).

Table 2. The Impact of the Coin Toss on Subsequent Behavior

	2 Months after Coin Toss			6 Months after Coin Toss		
	All	Important	Less Important	All	Important	Less Important
Heads	0.249*** (0.009)	0.111*** (0.012)	0.364*** (0.012)	0.211*** (0.012)	0.112*** (0.017)	0.295*** (0.016)
Prob of Change	0.445*** (0.017)	0.594*** (0.023)	0.279*** (0.024)	0.476*** (0.023)	0.597*** (0.032)	0.341*** (0.033)
Male	0.012 (0.009)	0.005 (0.013)	0.018 (0.012)	-0.001 (0.012)	-0.001 (0.018)	-0.003 (0.017)
Age	-0.002*** (0.001)	-0.003*** (0.001)	-0.002* (0.001)	-0.002** (0.001)	-0.006*** (0.001)	-0.001 (0.001)
Married	0.002 (0.011)	-0.014 (0.015)	0.013 (0.016)	-0.014 (0.015)	-0.037 (0.021)	0.015 (0.021)
US Resident	0.033*** (0.010)	0.020 (0.014)	0.039** (0.013)	0.008 (0.014)	0.002 (0.020)	0.007 (0.018)
Black	0.006 (0.027)	-0.029 (0.035)	0.041 (0.041)	-0.046 (0.039)	-0.127* (0.053)	0.042 (0.055)
Asian	0.004 (0.013)	0.008 (0.019)	-0.012 (0.018)	-0.022 (0.018)	-0.041 (0.027)	-0.023 (0.025)
Hispanic	0.019 (0.017)	0.007 (0.024)	0.027 (0.023)	-0.011 (0.023)	-0.013 (0.034)	-0.008 (0.032)
Race-Other	0.004 (0.022)	0.002 (0.031)	0.004 (0.030)	0.010 (0.032)	0.051 (0.046)	-0.038 (0.043)
4-Year College	-0.007 (0.010)	-0.013 (0.016)	-0.008 (0.014)	-0.003 (0.015)	0.008 (0.023)	-0.023 (0.019)
Income > 50K	0.000 (0.010)	-0.015 (0.014)	0.016 (0.014)	-0.008 (0.014)	-0.011 (0.020)	0.003 (0.019)
Live in a City	0.007 (0.009)	-0.003 (0.013)	0.013 (0.012)	-0.008 (0.012)	-0.003 (0.018)	-0.011 (0.016)
Pre-toss Happiness	-0.002 (0.002)	0.002 (0.003)	-0.003 (0.003)	-0.006 (0.003)	-0.004 (0.004)	-0.006 (0.005)
Best 2 of 3 Flip	-0.009 (0.011)	-0.006 (0.016)	-0.010 (0.014)	-0.013 (0.016)	-0.006 (0.024)	-0.019 (0.021)
Include Question Indicators	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10094	4607	5487	6131	2874	3257

Note: This table explores the impact of the coin toss on participants' subsequent behavior. Each column reports the results of a linear probability model in which the dependent variable is a dichotomous variable that corresponds to whether the survey respondent says a change was made. Columns 1 and 2 reflect two- and six-month survey responses, respectively, from the entire sample. Columns 3 and 4 present the same information for the subset of important questions, and Columns 5 and 6 correspond to the less important questions. Standard errors are reported in parentheses. *, **, *** denote significance at the 5, 1, and .1 percent levels.

Figure 7. Percentage Following the Coin Toss, Two-Month Survey



Note: Figure 7 presents the percentage of all respondents to the two-month survey who report taking the action that corresponds to the result of the coin toss. The questions are listed on the vertical axis and are divided into “important” and “unimportant” groupings. Questions with fewer than 150 responses were excluded from this figure.

Table 3. The Link between Choices and Self-reported Happiness

	Question	2 Months after Coin Toss			6 Months after Coin Toss		
		1st Stage	OLS	2SLS	1st Stage	OLS	2SLS
	All	0.249 (0.009)	0.449 (0.039)	0.041 (0.139)	0.211 (0.012)	0.584 (0.048)	0.476 (0.214)
	Important	0.111 (0.012)	0.782 (0.066)	0.554 (0.495)	0.112 (0.017)	1.011 (0.076)	2.153 (0.652)
	Unimportant	0.364 (0.012)	0.213 (0.047)	-0.073 (0.119)	0.295 (0.016)	0.190 (0.061)	-0.077 (0.194)
Important	Should I quit my job	0.059 (0.022)	1.643 (0.127)	0.905 (1.774)	0.070 (0.031)	1.890 (0.137)	5.203 (2.313)
	Should I break up	0.167 (0.030)	0.356 (0.159)	0.639 (0.818)	0.157 (0.040)	0.278 (0.192)	2.698 (1.259)
	Should I go back to school	0.119 (0.030)	0.595 (0.168)	-0.583 (1.162)	0.133 (0.042)	0.949 (0.190)	0.007 (1.280)
	Should I start my own business	0.168 (0.046)	0.399 (0.185)	0.000 (1.014)	0.077 (0.074)	0.520 (0.307)	5.256 (5.707)
	Should I move	0.004 (0.034)	0.795 (0.233)	56.326 (450.597)	0.087 (0.053)	0.823 (0.239)	3.176 (2.775)
	Should I quit smoking	0.129 (0.051)	0.160 (0.225)	1.417 (1.498)	0.147 (0.078)	0.313 (0.304)	-1.096 (1.995)
	Should I have a child	0.195 (0.046)	0.471 (0.261)	-1.598 (1.083)	0.193 (0.073)	0.395 (0.270)	-0.450 (1.288)
	Should I propose	0.183 (0.075)	0.362 (0.506)	1.021 (1.862)	-0.041 (0.124)	1.529 (0.640)	-5.125 (19.881)
Less Important	Should I splurge	0.303 (0.029)	0.197 (0.096)	-0.555 (0.312)	0.204 (0.035)	0.458 (0.136)	0.163 (0.594)
	Should I go on a diet	0.488 (0.032)	0.413 (0.126)	0.754 (0.252)	0.471 (0.044)	0.146 (0.176)	0.154 (0.361)
	Should I break my bad habit	0.607 (0.030)	0.146 (0.123)	-0.325 (0.199)	0.384 (0.044)	-0.001 (0.168)	-0.597 (0.420)
	Should I get a tattoo	0.111 (0.024)	0.524 (0.255)	-0.775 (1.270)	0.139 (0.044)	0.669 (0.260)	-1.123 (1.478)
	Should I try online dating	0.465 (0.044)	0.043 (0.180)	0.611 (0.377)	0.269 (0.060)	0.129 (0.249)	-0.429 (0.846)
	Should I join a gym	0.236 (0.045)	0.690 (0.188)	0.369 (0.686)	0.288 (0.065)	0.292 (0.209)	0.970 (0.671)
	Should I dye my hair	0.315 (0.052)	0.327 (0.188)	0.266 (0.553)	0.148 (0.068)	0.664 (0.261)	1.863 (1.623)
	Should I sign up for a running event	0.265 (0.049)	0.437 (0.192)	-0.790 (0.675)	0.347 (0.065)	-0.234 (0.216)	-0.395 (0.572)
	Should I grow facial hair	0.390 (0.053)	-0.137 (0.209)	-0.275 (0.467)	0.234 (0.079)	-0.726 (0.334)	-0.624 (1.149)
	Should I quit drinking	0.446 (0.059)	-0.309 (0.246)	-0.427 (0.507)	0.278 (0.087)	-0.083 (0.316)	1.150 (1.068)
	Should I ask for a raise	0.356 (0.064)	-0.037 (0.276)	-0.689 (0.712)	0.425 (0.087)	0.115 (0.375)	-1.116 (0.833)
	Should I start volunteering	0.303 (0.054)	0.037 (0.274)	-0.135 (0.714)	0.478 (0.071)	0.090 (0.303)	0.129 (0.510)
	Observations	10094	10094	10094	6131	6131	6131

Note: This table presents regression results exploring the link between choices and self-reported happiness. Columns 1 to 3 correspond to the two-month survey; Columns 4 to 6 correspond to the six-month survey. Columns 1 and 4 are first-stage estimates and describe the degree to which the coin toss affected the action taken. Columns 2 and 5 are OLS estimates, which show the extent to which those who make a change are more or less happy than those who maintain the status quo. Columns 3 and 6 are the instrumental variable estimates. The row “Observations” reflects the number of observations in the regression that includes all questions. Questions with fewer than 150 respondents were included in the first panel but are not presented as separate regressions. Standard errors are reported in parentheses.

Table 4. Sensitivity Analysis for Important Questions

Question	2 Months after Coin Toss				6 Months after Coin Toss			
	1st Stage	OLS	2SLS	Observations	1st Stage	OLS	2SLS	Observations
All	0.111 (0.012)	0.782 (0.066)	0.554 (0.495)	4607	0.112 (0.017)	1.011 (0.076)	2.153 (0.652)	2874
Female	0.129 (0.018)	0.839 (0.103)	0.953 (0.660)	2020	0.121 (0.026)	1.030 (0.111)	2.119 (0.877)	1282
Male	0.096 (0.016)	0.742 (0.085)	0.158 (0.753)	2587	0.103 (0.023)	0.997 (0.105)	2.256 (0.977)	1592
Younger than 30	0.128 (0.018)	0.604 (0.088)	0.379 (0.596)	2337	0.119 (0.025)	0.808 (0.103)	2.332 (0.868)	1429
30 or Older	0.091 (0.017)	0.959 (0.098)	0.739 (0.868)	2270	0.107 (0.024)	1.204 (0.112)	2.148 (0.989)	1445
No Friend Named	0.107 (0.015)	0.695 (0.080)	0.850 (0.626)	3097	0.115 (0.021)	0.916 (0.093)	1.548 (0.755)	1892
Friend Named	0.121 (0.022)	0.938 (0.114)	-0.021 (0.802)	1510	0.105 (0.030)	1.218 (0.133)	3.233 (1.287)	982
Income Below 50K	0.096 (0.018)	0.741 (0.093)	-0.771 (0.873)	2231	0.091 (0.026)	0.820 (0.111)	0.573 (1.121)	1356
Income Above 50K	0.126 (0.017)	0.796 (0.092)	1.459 (0.607)	2376	0.131 (0.024)	1.167 (0.106)	3.118 (0.821)	1518
Report Unlikely to Follow Toss	0.061 (0.016)	0.772 (0.091)	-0.716 (1.261)	2584	0.064 (0.023)	0.948 (0.104)	3.324 (1.702)	1609
Report Likely to Follow Toss	0.170 (0.019)	0.762 (0.096)	1.027 (0.490)	2011	0.172 (0.027)	1.076 (0.114)	1.523 (0.629)	1256
Below Average Pre-Toss Happiness	0.127 (0.021)	1.105 (0.123)	1.088 (0.810)	1692	0.151 (0.029)	1.417 (0.141)	2.161 (0.876)	1045
Above Average Pre-Toss Happiness	0.102 (0.015)	0.578 (0.075)	0.143 (0.609)	2915	0.092 (0.022)	0.754 (0.087)	1.967 (0.917)	1829
Uncertain about Changing Pre-Toss	0.084 (0.016)	0.694 (0.103)	-0.530 (0.963)	2306	0.092 (0.024)	0.956 (0.113)	1.845 (1.122)	1429
(Un)likely to Change Pre-Toss	0.137 (0.019)	0.840 (0.086)	1.163 (0.561)	2280	0.133 (0.025)	1.091 (0.104)	2.662 (0.797)	1430

Note: This table presents a sensitivity analysis for the important questions. Columns 1 to 3 correspond to the two-month survey; Columns 4-6 correspond to the six-month survey. Columns 1 and 4 are first-stage estimates and describe the degree to which the coin toss affected the action taken. Columns 2 and 5 are OLS estimates, which show the extent to which those who make a change are more or less happy than those who maintain the status quo. Columns 3 and 6 are the instrumental variable estimates. The top row of Table 4 replicates the second row of Table 3, which serves as the baseline specification against which the other results of Table 4 can be compared. The remaining rows categorize the participants by gender, age, and the like and evaluate the robustness of the results presented in Table 3. Standard errors are reported in parentheses.

Table 5. Are Tossers Who Follow the Toss More Likely to Report?

	Third Party Says Tosser Followed Toss	Third Party Says Tosser Did Not Follow Toss	Difference
Two-Month Survey	0.860	0.799	0.061
Six-Month Survey	0.747	0.681	0.066

Note: This table explores whether the survey response rate for important questions is affected by whether the tosser follows the result of the flip. Columns 1 and 2 present tosser response rates according to whether the third party reported that the tosser did or did not follow the toss. Column 3 reports the resulting difference between the first two columns. The rows divide the results by two- and six-month survey responses.

Table 6. Are Happy Changers Especially Likely to Report?

	Third Party Says Tosser Made a Change	Third Party Says Tosser Did Not Make a Change	Difference
Two-Month Survey			
Third Party Says Tosser Is Happier than Average	0.897	0.837	0.060
Third Party Says Tosser Is Less Happy than Average	0.822	0.810	0.012
Six-Month Survey			
Third Party Says Tosser Is Happier than Average	0.828	0.683	0.144
Third Party Says Tosser Is Less Happy than Average	0.677	0.637	0.040

Note: This table explores whether the survey response rate for important questions is higher among happy tossers who make a change. The percent of tossers who completed a survey is presented in the cells. The first two columns divide responses according to whether the third party reported that the tosser made a change. The third column takes the difference between the first two columns. Rows divide the sample by whether the third party reported that the tosser's happiness was above- or below- average. The two panels reflect the two- and six-month survey responses, respectively.

Table 7. Are Happy Heads and Sad Tails Especially Likely to Report?

	Heads Result	Tails Result	Difference
Two-Month Survey			
Third Party Says Tosser Is Happier than Average	0.849	0.864	-0.015
Third Party Says Tosser Is Less Happy than Average	0.794	0.823	-0.029
Six-Month Survey			
Third Party Says Tosser Is Happier than Average	0.771	0.752	0.019
Third Party Says Tosser Is Less Happy than Average	0.663	0.630	0.033

Note: This table explores whether the survey response rate for important questions is higher among happy tossers who flip heads and sad tossers who flip tails. The percent of tossers who completed a survey is presented in the cells. The first two columns divide responses according to whether the tosser flipped heads versus tails. The third column takes the difference between the first two. Rows divide the sample by whether the third party reported that the tosser's happiness was above- or below- average. The two panels reflect the two- and six-month survey responses, respectively.

Table 8. Do Tossers Claim to Have Followed the Toss When They Have Not Actually Done So?
Conditional on Having a Response from Both the Tosser and Third Party

	Tosser	Third Party
Two-Month Survey	0.614	0.580
Six-Month Survey	0.550	0.557

Note: This table explores whether tossers overreport having followed the toss for important questions. Column 1 presents the rate at which tossers report following the toss, while Column 2 presents the same information based on third party reports. The rows show results from the two- and six-month surveys, respectively.

Table 9. Do Participants Who Make a Change Exaggerate How Happy They Are?

	OLS 2M	OLS 6M
Tosser Report of Own Happiness	0.449 (0.039)	0.584 (0.048)
Tosser Report of Own Happiness Conditional on Having Third Party Response	0.472 (0.083)	0.635 (0.131)
Third Party Report of Tosser Happiness Conditional on Having Tosser Response	0.380 (0.095)	0.640 (0.149)

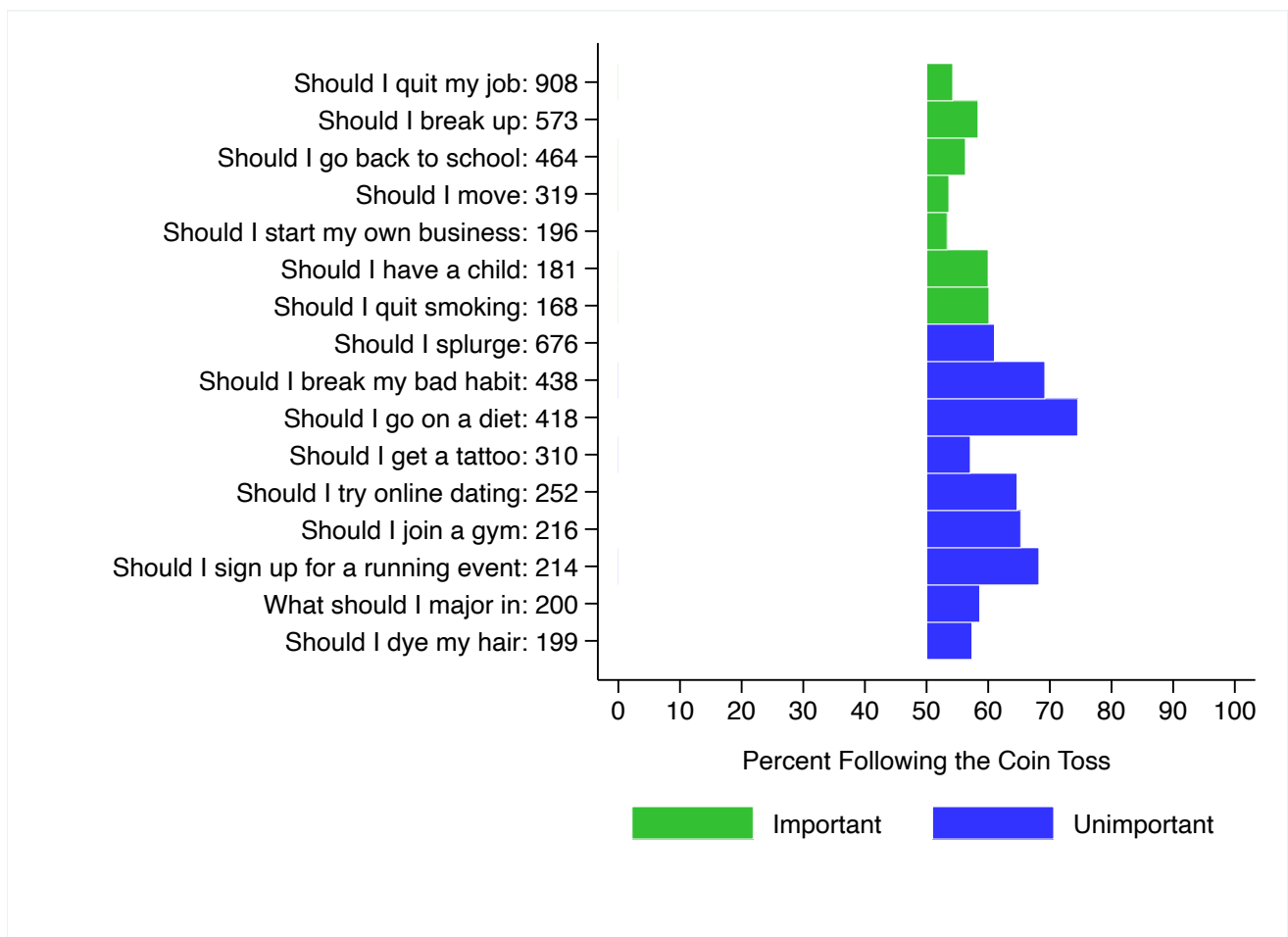
Note: This table explores whether tossers who made a change are likely to exaggerate how happy they are. The first row presents the coefficient on whether the individual made a change from OLS regressions with the tosser's self-reported happiness as the lefthand variable. The second row presents the same information but conditional on having a response from the third party. The third row replaces the lefthand variable with the third party's report of the tosser's happiness. Columns report OLS results by two- and six-month survey results. Standard errors are reported in parentheses.

Appendix Table 1. Sensitivity Analysis for Unimportant Questions

Question	2 Months after Coin Toss				6 Months after Coin Toss			
	1st Stage	OLS	2SLS	Observations	1st Stage	OLS	2SLS	Observations
All	0.364 (0.012)	0.213 (0.047)	-0.073 (0.119)	5487	0.295 (0.016)	0.190 (0.061)	-0.077 (0.194)	3257
Female	0.370 (0.018)	0.318 (0.073)	0.117 (0.182)	2380	0.288 (0.024)	0.282 (0.092)	0.358 (0.299)	1415
Male	0.362 (0.015)	0.133 (0.061)	-0.199 (0.155)	3107	0.296 (0.020)	0.077 (0.083)	-0.406 (0.258)	1842
Younger than 30	0.359 (0.015)	0.151 (0.059)	-0.117 (0.152)	3440	0.280 (0.020)	0.179 (0.077)	0.077 (0.253)	2040
30 or Older	0.370 (0.019)	0.306 (0.077)	-0.013 (0.190)	2047	0.317 (0.026)	0.204 (0.104)	-0.262 (0.305)	1217
No Friend Named	0.314 (0.015)	0.206 (0.064)	-0.026 (0.187)	3271	0.252 (0.021)	0.196 (0.084)	0.372 (0.308)	1860
Friend Named	0.439 (0.018)	0.226 (0.067)	-0.107 (0.142)	2216	0.353 (0.024)	0.186 (0.090)	-0.446 (0.242)	1397
Income Below 50K	0.361 (0.015)	0.207 (0.063)	-0.053 (0.160)	3273	0.278 (0.021)	0.169 (0.080)	0.171 (0.269)	1933
Income Above 50K	0.367 (0.018)	0.220 (0.069)	-0.140 (0.175)	2214	0.315 (0.024)	0.231 (0.097)	-0.394 (0.286)	1324
Report Unlikely to Follow Toss	0.167 (0.023)	0.255 (0.105)	0.272 (0.536)	1363	0.063 (0.032)	0.437 (0.131)	-0.644 (1.908)	811
Report Likely to Follow Toss	0.427 (0.013)	0.191 (0.052)	-0.135 (0.115)	4114	0.369 (0.018)	0.122 (0.070)	-0.035 (0.178)	2442
Below Average Pre-Toss Happiness	0.346 (0.020)	0.242 (0.091)	0.050 (0.242)	1967	0.234 (0.028)	0.394 (0.122)	0.124 (0.482)	1120
Above Average Pre-Toss Happiness	0.377 (0.014)	0.192 (0.052)	-0.144 (0.128)	3520	0.331 (0.019)	0.082 (0.069)	-0.108 (0.193)	2137
Uncertain about Changing Pre-Toss	0.341 (0.018)	0.204 (0.075)	0.014 (0.200)	2198	0.235 (0.025)	0.262 (0.104)	0.154 (0.395)	1292
(Un)likely to Change Pre-Toss	0.379 (0.015)	0.221 (0.060)	-0.136 (0.147)	3263	0.334 (0.020)	0.131 (0.077)	-0.207 (0.218)	1953

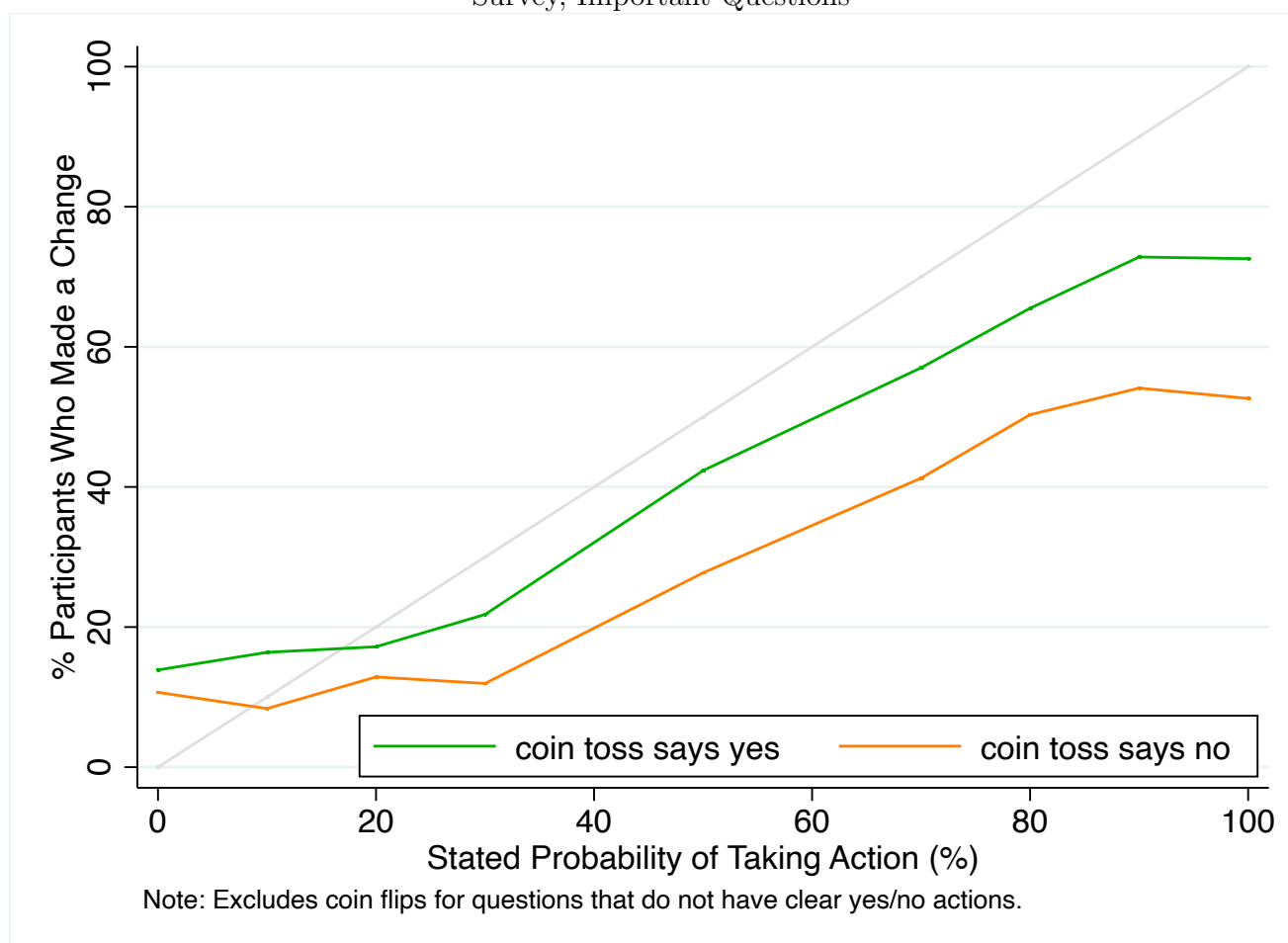
Note: This table presents a sensitivity analysis for the unimportant questions. Columns 1 to 3 correspond to the two-month survey; Columns 4 to 6 correspond to the six-month survey. Columns 1 and 4 are first-stage estimates and describe the degree to which the coin toss affected the action taken. Columns 2 and 5 are OLS estimates, which show the extent to which those who make a change are more or less happy than those who maintain the status quo. Columns 3 and 6 are the instrumental variable estimates. The top row of Appendix Table 1 replicates the third row of Table 3, which serves as the baseline specification against which the other results of Appendix Table 1 can be compared. The remaining rows categorize the participants by gender, age, and the like and evaluate the robustness of the results presented in Table 3. Standard errors are reported in parentheses.

Appendix Figure 1. Percentage Following the Coin Toss, Six-Month Survey



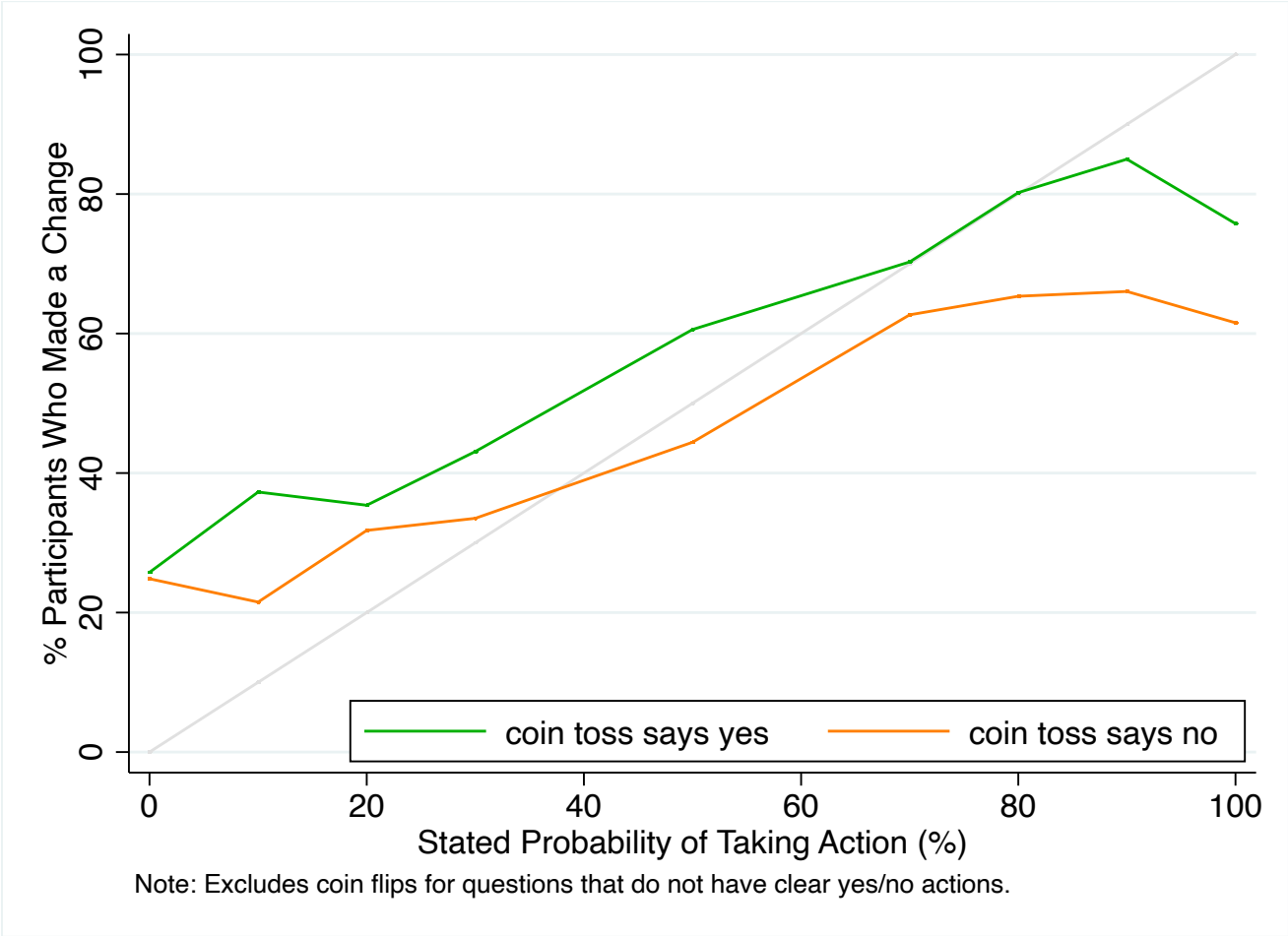
Note: Appendix 1 presents the percentage of all respondents to the six-month survey who report taking the action that corresponds to the result of the coin toss. The questions are listed on the vertical axis and are divided into “important” and “unimportant” groupings. Questions with fewer than 150 responses were excluded from this figure.

Appendix Figure 2. Likelihood of Taking Action as a Function of Ex Ante Stated Probabilities, Two-Month Survey, Important Questions



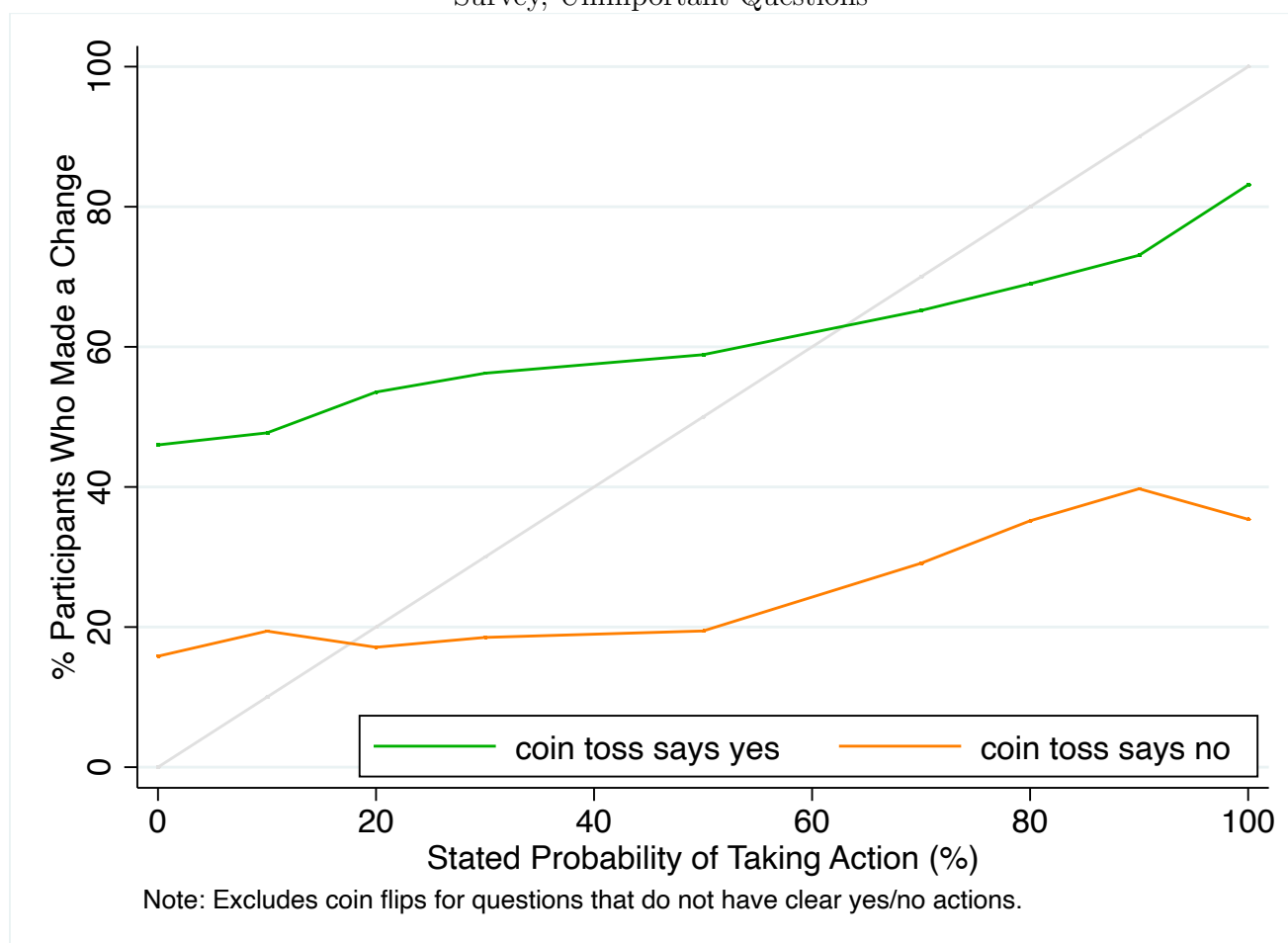
Note: Appendix Figure 2 presents the percent of participants who make a change by the two-month survey mark and flipped the coin on an important question according to their stated probability of changing and the result of the coin flip. The vertical axis reflects the percent of respondents who reported making a change. The horizontal axis groups respondents according to to their stated ex-ante likelihoods of making a change. Responses are categorized according to whether the coin came up heads (make a change) or tails (no change).

Appendix Figure 3. Likelihood of Taking Action as a Function of Ex Ante Stated Probabilities, Six-Month Survey, Important Questions



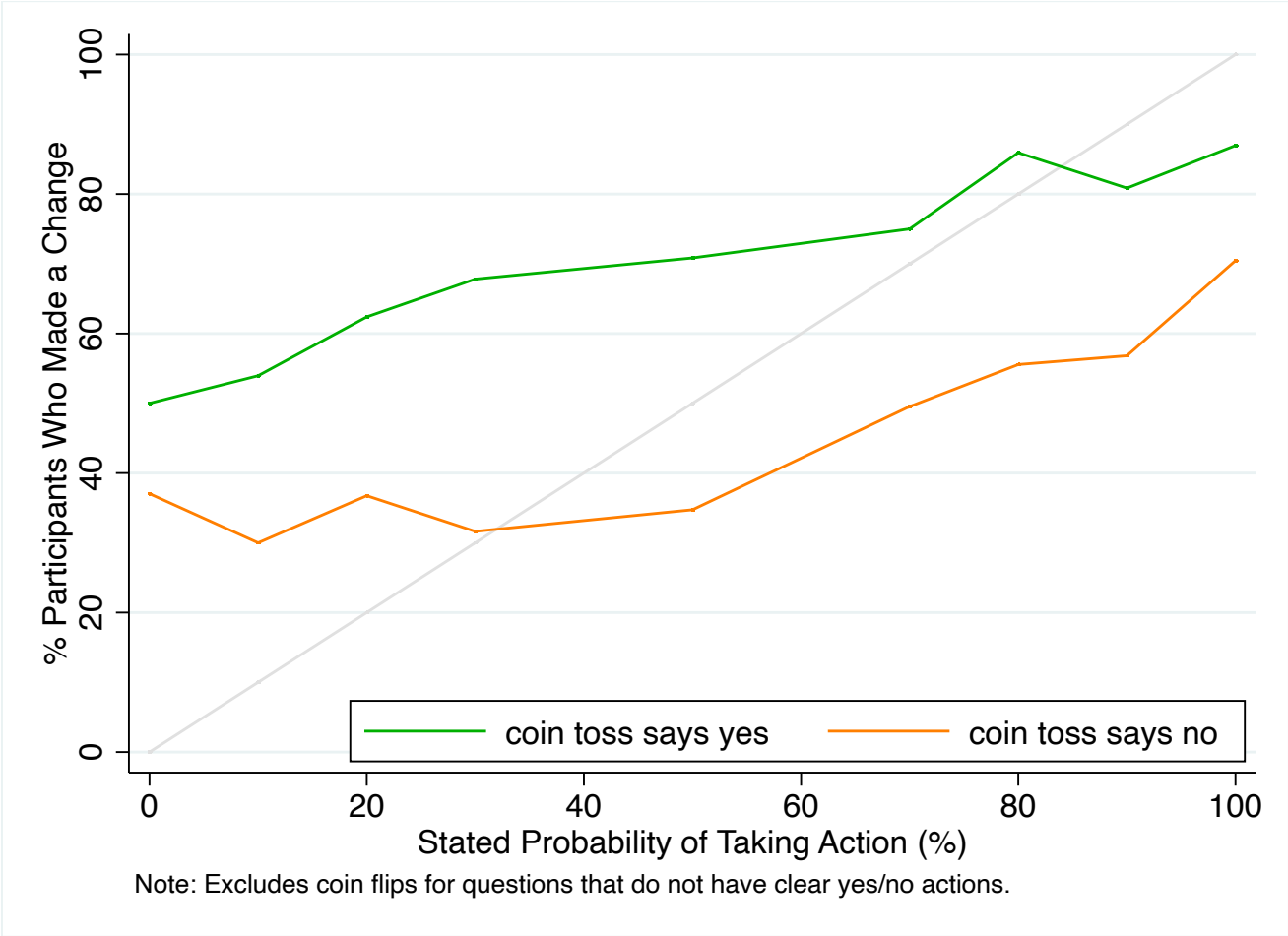
Note: Appendix Figure 3 presents the percent of participants who make a change by the six-month survey mark and flipped the coin on an important question according to their stated probability of changing and the result of the coin flip. The vertical axis reflects the percent of respondents who reported making a change. The horizontal axis groups respondents according to to their stated ex-ante likelihoods of making a change. Responses are categorized according to whether the coin came up heads (make a change) or tails (no change).

Appendix Figure 4. Likelihood of Taking Action as a Function of Ex Ante Stated Probabilities, Two-Month Survey, Unimportant Questions



Note: Appendix Figure 4 presents the percent of participants who make a change by the two-month survey mark and flipped the coin on an unimportant question according to their stated probability of changing and the result of the coin flip. The vertical axis reflects the percent of respondents who reported making a change. The horizontal axis groups respondents according to their stated ex-ante likelihoods of making a change. Responses are categorized according to whether the coin came up heads (make a change) or tails (no change).

Appendix Figure 5. Likelihood of Taking Action as a Function of Ex Ante Stated Probabilities, Six-Month Survey, Unimportant Questions



Note: Appendix Figure 5 presents the percent of participants who make a change by the six-month survey mark and flipped the coin on an unimportant question according to their stated probability of changing and the result of the coin flip. The vertical axis reflects the percent of respondents who reported making a change. The horizontal axis groups respondents according to to their stated ex-ante likelihoods of making a change. Responses are categorized according to whether the coin came up heads (make a change) or tails (no change).