NBER WORKING PAPER SERIES

CHOOSING YOUR POND: LOCATION CHOICES AND RELATIVE INCOME

Nicolas L. Bottan
Ricardo Perez-Truglia

Working Paper 23615
http://www.nber.org/papers/w23615

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2017, Revised December 2017

Previously circulated as "Choosing Your Pond: Revealed-Preference Estimates of Relative Income Concerns" and "Choosing Your Pond: Measuring Preferences for Relative Consumption." We are thankful for comments from David Albouy, Dan Bernhardt, Jeff Brown, Leonardo Bursztyn, Mikhail Galashin, Craig Fox, Ori Heffetz, Erzo F.P. Luttmer, Ben Marx, Alex Rees-Jones, Chris Roth, Melanie Wasserman and seminar discussants at UCLA-Anderson, UIUC, Loyola Marymount University, 2017 SITE Workshop, 2017 Advances with Field Experiments Conference, 2017 NTA Annual Conference on Taxation, 2017 SEA Annual Meeting, 2017 GEM-BPP Workshop, Binghamton University, the Federal Reserve Bank of New York and the University of Pennsylvania. We thank UCLA-Anderson and its Behavioral Lab for providing funding for the experiment; and thank the support from the Robert Ferber Dissertation Award. This project was reviewed and approved in advance by the Institutional Review Board at University of California Los Angeles (IRB #16-001968; #17-001449). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Choosing Your Pond: Location Choices and Relative Income
Nicolas L. Bottan and Ricardo Perez-Truglia
NBER Working Paper No. 23615
July 2017, Revised December 2017
JEL No. D62,D91,I31,Z13

ABSTRACT

We provide unique revealed-preference evidence that, when choosing where to live, individuals care about their position in the income distribution. We study the decisions of senior medical students in the National Resident Matching Program (NRMP). They must choose between programs that offer similar nominal incomes, but in cities with different costs of living and income distributions. We conduct a survey experiment with 1,100 NRMP participants to elicit their perceptions about cost of living and relative income in their prospective cities and their rank order submissions. To assess the direction of causality, we embed an information-provision experiment that generates exogenous variations in perceived cost of living and relative income. We find evidence that, in addition to the cost of living, individuals care about their relative income. Moreover, we find substantial and meaningful heterogeneity by relationship status in preferences for relative income. We conduct a complementary survey experiment to assess the robustness of our results and to disentangle confounding factors. The evidence is consistent with a combination of relative concerns and dating expectations.

Nicolas L. Bottan
Department of Economics
University of Illinois at Urbana-Champaign
1407 W Gregory Dr.
214 David Kinley Hall, Mc-707
Urbana, IL 61801
bottan2@illinois.edu

Ricardo Perez-Truglia
Anderson School of Management
University of California, Los Angeles
110 Westwood Plaza
Los Angeles, CA 90095
and NBER
ricardo.truglia@anderson.ucla.edu

1 Introduction

It is well documented that individuals care about and respond to differences in cost of living when deciding where to live. Many theories predict that, holding the cost of living constant, individuals also care about their rank in the income distribution. For instance, relative income may matter for social interactions, such as dating outcomes (Fisman et al. 2006) and eliciting envy and pride (Luttmer 2005). In this paper, we provide the first test of this hypothesis that relative income influences individuals' location decisions by means of a field experiment with 1,100 medical students in the National Resident Matching Program (NRMP).

To study how individuals trade off between cost of living and relative income, it would be useful to exploit data on how individuals choose from a list of cities that offer different combinations of cost of living and relative income. This setting would allow comparisons of how individuals trade off between these two attributes. However, such datasets do not exist. Although some datasets identify individuals moving from one location to another,[1] they do not include sufficient information to estimate preferences. For instance, they do not identify alternative combinations of cost of living and relative income that the individual could have gotten in other locations. Even if such data were available, one would have to address omitted variable biases.

In this paper, although we do not collect the ideal dataset, we get closer to it than any previous study. We accomplish this by exploiting a unique context, the NRMP, which uses an algorithm to pair medical students in the United States with hospital residency programs, based on rankings submitted by students and hospitals. Students submit their rankings to the NRMP, and the resulting match determines where they will live for roughly five years. They choose between programs that offer almost identical nominal incomes, but in cities with largely different costs of living and income distributions. Thus, medical students face substantial tradeoffs between cost of living and relative income.

Several features of the NRMP make it desirable for this type of revealed-preference analysis (Benjamin et al. 2014). First, the deadline for students to submit their rankings creates an identifiable moment when the decision becomes irreversible. Second, it is possible to identify and specify the entire choice set faced by these individuals. Third, because most students are aware of the incentive-compatible matching algorithm used by the NRMP, it is possible to infer preferences directly without estimating a model that relies on additional assumptions. Fourth, this is a high-stakes choice to which participants devote ample time and attention: medical students devote their entire fourth years of medical school to the Match and two months to ranking residency programs. This decision is arguably one of the most important

[1] For instance, the National Survey of Families and Households (Luttmer 2005), or the United States Postal Service's National Change of Address (Perez-Truglia 2017).

for their careers and lives.

We conducted a survey of 1,100 senior medical students participating in the 2017 Main Residency Match. The survey asked participants to list their top two favorite residency programs, because these two choices have the highest stakes and receive the most individual attention. We elicited perceptions about two aspects of the cities in which these two programs are located: the cost of living and earnings rank (i.e., the position in the distribution of individual earnings in the city).

We then elicited each subject's expected rank submission. Using these data on perceptions and choices, we can estimate how differences in the cities' costs of living and relative incomes affect location choices. On the one hand, we expect individuals to prefer low costs of living, which means high consumption. On the other hand, individuals also may prefer high or low earnings rank, depending on certain mechanisms. For example, models of relative concerns like envy and pride predict that individuals want to increase their rank in their reference group, so they prefer less affluent ponds. On the contrary, single individuals looking for a rich partner may prefer more affluent ponds.

Perceptions about cost of living and earnings rank may be correlated with unobservable attributes of the alternatives, which can generate omitted-variable biases. To deal with this concern, we embedded an information-provision experiment in the survey. Immediately after eliciting perceptions about cost of living and earnings rankings, we provided all individuals with statistics about these two measures. We randomized the value of this feedback in a non-deceptive way by randomizing the data source used to compute these statistics. For instance, students considering earning $54,000 at a residency in Champaign-Urbana, IL, received one of two messages: their earnings rank would be 55.1% according to data from the Current Population Survey or 60.3% according to data from the American Community Survey. We then elicited their perceptions again after providing this feedback. This source-randomization experiment created exogenous variation in posterior beliefs. We exploited that exogenous variation in an instrumental variables model to estimate the causal effects of the perceived attributes, cost of living, and earnings rank on choice.

Our baseline estimates, which use experimental and non-experimental variations in perceptions, suggest that a 1 percentage point decrease in cost of living increases the probability that a program is chosen by 0.201 percentage points (i.e., a behavioral elasticity of 0.201). This concern for affordability is consistent with the fact that half of pre-med students report money to be the primary motivation for their career choices (Daniel and O'Brien 2008). Although this preference for cost of living is statistically and economically significant, it is by no means the primary concern for medical students. Using perceptions about other characteristics of the residencies, we find that doctors care substantially more about prestige and career prospects than about cost of living.

Most important, our baseline estimates suggest that, holding cost of living constant, individuals also care about relative income: on average, an increase of 1 percentage point in earnings rank increases the probability that a program is chosen by 0.186 percentage points (i.e., a behavioral elasticity of 0.186). These baseline estimates suggest that the average individual gives roughly the same importance to cost of living as to relative income. However, these average preferences mask meaningful heterogeneity. Non-single individuals (i.e., those who are married or in a long-term relationship) prefer locating to less affluent ponds. On the contrary, single individuals prefer locating to more affluent ponds. This difference in preferences is large in magnitude and highly statistically significant (p-value < 0.001). The heterogeneity is consistent with prior evidence from the happiness literature: Luttmer (2005) finds that the positive effects of relative income on happiness are driven entirely by non-singles individuals. Also, the direction of this heterogeneity is consistent with prior evidence that singles prefer to date rich partners (Fisman et al. 2006) and prefer a more dense dating market (Gautier et al. (2010)) and thus should be attracted to more affluent cities.

We find that the baseline estimates of preferences for relative income satisfy several robustness checks. First, these estimates are not sensitive to the inclusion of multiple residency and location characteristics as control variables. Second, they are consistent with the experimental estimates that focus on the variation generated by the information provision experiment.[2] Third, we find that the information related to earnings rank, which is provided roughly one month before the submission deadline, has a lasting effect on the final rank order submitted to the NRMP.

We conduct an additional robustness check consisting of an auxiliary experiment based on an online sample of respondents recruited through Amazon Mechanical Turk. On the one hand, this sample has a disadvantage: respondents are not planning to move anytime soon, so we cannot measure the effects on their actual location choices. Thus, we instead measure the effects of the information provision on their hypothetical location choices. On the other hand, the auxiliary sample has a couple of advantages. Most important, due to wider availability of subjects, it is possible to run additional experiments at any time. Additionally, the set of online respondents is more diverse than the medical students in many dimensions, such as age and occupation, allowing us to assess the external validity of the results.

The findings from this auxiliary experiment are consistent with and more precisely estimated than the findings from the main experiment: we find that the average individual prefers a high relative income and that this preference is stronger for non-singles than it is

[2]The baseline estimates of preferences for cost of living are smaller in magnitude in these specifications. Thus, if anything, our baseline estimates underestimate the importance of relative income.

for singles. Moreover, despite large differences in observable characteristics between the two samples, the findings are quantitatively similar: the marginal rate of substitution between relative income and cost of living is 0.90 (s.e. 0.64) in the main experiment and 0.66 (s.e. 0.20) in the auxiliary experiment, and the difference is statistically insignificant. This evidence suggests that preferences for relative income are not exclusive to the sample of medical students.

In the last part of the paper, we provide evidence about the potential mechanisms at play. One potential concern is that individuals do not care about their relative income *per se* but that they use it as a signal for other unobservable attributes of the city. We think this possibility is unlikely in this environment of high stakes and high information. Medical students devote two months to this decision, during which they travel to the locations under consideration and they gather and analyze a lot of information about the residency programs and their locations. It is unlikely that they would need information about earnings rank to learn about other features of the locations and programs, because they can learn this information directly. Nevertheless, we provide a direct test of this confounding factor.

One version of this confounding factor is that individuals react to information about relative income because it teaches them about the expected cost of living. If the students expect to compete with their neighbors for some goods like housing, then it would be natural for them to make such inferences. Contrary to this hypothesis, we show that feedback about earnings rank has a zero and precisely estimated effect on expected cost of living, both in the short term (the baseline survey) and in the long term (the follow-up survey).

A second version of this confounding factor is that individuals react to information about relative income because it teaches them about location characteristics like public goods and crime rates. Although such inferences are not unnatural, they operate in the opposite direction: because more affluent cities tend to have more desirable amenities than less affluent cities, this mechanism predicts that individuals will prefer more affluent ponds, which is the opposite of what we find. To address this confounding factor, we extend the survey instrument used in the auxiliary sample so that, after individuals received feedback about relative income, we elicit perceptions about other city characteristics, such as school quality and crime rates. We find evidence that is consistent with the previous argument: if anything, this confounding factor leads to a slight underestimation of the preferences for relative income.

Considering this additional evidence, our favorite interpretation for the estimated preference for relative income among non-singles is that they care about relative income *per se*. The literature on relative concerns (Luttmer 2005) provides several explanations for this preference. Individuals may anticipate that their consumption aspirations will increase with their neighbors' consumption (Frank 1985). Individuals could feel happier from observing that they do better than their neighbors (Luttmer 2005). Also, individuals may expect to be treated better by peers when they rank higher in the income distribution (Doob and Gross 1968; Fennis 2008; Nelissen and Meijers 2011). Regarding the heterogeneity in preferences for relative income by relationship status, our favorite interpretation consists of dating prospects: being single increases the attractiveness of more affluent ponds, because more affluent individuals are more desirable partners than less affluent individuals (Fisman et al. 2006).

Our paper is related to several bodies of research, including a literature documenting that individuals have substantial misperceptions of their own income rank (e.g., Cruces et al. 2013; Karadja et al. 2017). This literature shows that correcting these misperceptions has significant effects on stated preferences for redistribution. Yet, there is no evidence that these misperceptions have a significant effect on behavior. We fill this gap in the literature by showing that misperceptions about relative income can have meaningful economic consequences.

Our study relates to literature on the importance of relative income for subjective well-being. Since the seminal contribution by Easterlin (1974), several studies have argued that, with own income held constant, subjective well-being increases with the relative income in the area of residence (Van de Stadt et al. 1985; Clark and Oswald 1996; Luttmer 2005; Ferrer-i Carbonell 2005; Perez-Truglia 2016).[3] Our estimates are not directly comparable to those from the happiness literature because of differences in contexts and in model specifications. With that caveat in mind, our baseline estimates suggest that relative income concerns are smaller in magnitude than the relative concerns estimated with happiness data by Luttmer (2005). However, due to lack of precision, we cannot reject the hypothesis that these two estimates are equal. Additionally, our results highlight an important aspect that is impossible to study with happiness data: even if happiness depends on relative income, it is unclear whether individuals anticipate the externalities brought by neighbors (Luttmer 2005). Our findings suggest that individuals anticipate these externalities, at least partially.

This paper also is related to studies using surveys, in which respondents must choose between pairs of hypothetical scenarios that encompass tradeoffs between income and status or between positional and non-positional goods. These studies find that individuals are sometimes willing to exchange absolute income for higher status (e.g., Solnick and Hemenway (1998); Johansson-Stenman et al. (2002); Yamada and Sato (2016); Andrew E. Clark and

[3]These studies often use a slightly different specification: holding own income constant, well-being decreases with the average income in the group of reference. It must be noted that some studies find the opposite effect (Senik 2004) or mixed evidence (Clark et al. 2009). For an extensive review of the literature, see Tideman et al. 2008.

Yamada (2017)). Similarly, evidence from laboratory experiments shows that relative standing affects the decision to buy a risky asset or share with others (e.g., Kuziemko et al. 2014). We contribute to this literature by estimating these tradeoffs in a real-world, high-stakes context.

Our study also contributes to literature that studies conspicuous consumption using observational data (Heffetz 2011) and experimental data (Bursztyn et al. 2017a; Roth 2015). Last, this paper relates to a literature that uses a variety of methods to measure how location amenities and cost of living influence household location decisions (e.g., Albouy 2008).

The rest of the paper proceeds as follows. Section 2 describes the survey design. Section 3 presents the econometric model. Section 4 presents implementation details and descriptive statistics. Section 5 discusses the distribution of perceptions and learning. Section 6 presents the preferences for cost of living and relative income. Section 7 presents the results from the auxiliary survey experiment. Section 8 presents supporting evidence for the interpretation of our results. The last section concludes.

2 Survey Design

2.1 Timing of the Surveys

After graduating from medical school, students have to complete a residency to become a Medicinae Doctor (MD). A residency usually lasts from three to seven years, after which individuals may obtain their medical license. During the fall semester of 2016, fourth year medical school students started their participation in the residency match by submitting applications to residency programs. Later in the semester, they were interviewed and flown out by some of the programs they applied to.[4] After all interviews were completed, the students spend almost two months deciding how to rank their favorite programs. Students have visited the cities during their interviews, and sometimes visit them again during this period. During this time, deciding on the rank order preference is the students' top priority: applicants claim to collect a lot of information to aid their decision, such as characteristics of the residency programs and characteristics of the cities where the programs are located.

We follow Benjamin et al. (2014) in using this context to study preferences. They conducted a survey of medical students after the students had submitted their rankings to the NRMP. The survey measured the submitted rankings as well as the perceived rank of many aspects of the programs, such as life satisfaction, happiness, and sense of control. In

[4] In 2015, the median number of applications submitted was 30 and the median number of interviews 16 (NRMP 2015).

that study, Benjamin et al. (2014) measure and compare the preferences inferred from rank choices to those inferred from subjective well-being. We follow the survey collection method from Benjamin et al. (2014) closely, but we change the survey itself to test a different hypothesis, that is, whether individuals make a trade-off between relative income and cost of living. In doing so, we deviate from the survey design in an important aspect: we collect our baseline survey *before* subjects submit their rank choices to the NRMP. We also embed an information-provision experiment, which allows us to address concerns on causality.

In the 2017 Match, the submission window for rank order lists opened on January 15 and closed on February 22. We conducted a baseline survey early in the submission period, which we describe first. We also conducted a follow-up survey after the submission window closed, which is described in Section 2.3.

2.2 General Structure of Baseline Survey

The baseline survey starts and ends with some background questions, such as the subject's medical school and marital status (see Appendix A.1 for the full questionnaire of the baseline survey). The core of the survey comprises the following group of questions, in the order listed below:

1. Choice Set: Elicit the names of the two favorite programs that the individual was considering for his or her order rank submission.

2. Prior Beliefs: Elicit perceptions about the cost of living and the earnings rank in the cities where these two programs are located.

3. Feedback: Provide subjects with feedback related to their perceptions.

4. Posterior Beliefs: Re-elicit perceptions about the cost of living and the earnings rank.

5. Rank Choice: Elicit the individual's expected rank submission (between the two programs).

The following sections provide details about each of these modules.

2.2.1 Choice Set

The survey asks individuals to list their top two preferred programs, in no particular order, from a user-friendly list of all the available programs organized by state and metro area. We limited the survey to two programs because otherwise it would have been too cognitively demanding. Most participants expect to be matched to one of their top-two choices: similar to previous years, 50.9% of the participants in the 2017 match were assigned to their first choice and 16.6% were assigned to their second choice. We concentrated on the participants' top two programs rather than a random pair of programs because this happen to be the part of the decision with the highest stakes and to which individuals were paying the most attention. In any case, our focus on the top two choices does not challenge the validity of our estimates: the research design would be valid with any pair (or group) of options, not only the top-two.[5]

2.2.2 Perceptions about Cost of Living and Earnings Rank

One important feature of the residency match process is that salaries are relatively homogeneous across the different programs, even across specialties.[6] Indeed, each program offers the same salary to all its candidates (and that salary is often publicly available on the program's website). Despite the homogeneity in nominal incomes, there is large heterogeneity in costs of living and earnings distributions in the cities where the programs are located. When designing the survey, we were constrained to using metropolitan areas rather than other geographical levels of aggregation (e.g., commuting zones) because the sources of data on cost of living are not collected at a finer level than the metro area.

We asked two questions about cost of living (one for each metro area) and two questions about the earnings rank (one for each metro area), in that order. For the cost of living question, we provided the following brief introduction: "You probably noticed that the average prices of goods and services are different across different cities. As a result, with the same income, you would be able to buy more things in some cities and less in other cities." After this introduction, we asked individual how much more or less expensive each metro area was, relative to the U.S. average. To make answering the question easier, we split it in two questions. The first question was: "Imagine that you chose to work in the [Metro Name] metro area. Would you expect your cost of living in this city to be cheaper or more expensive than the U.S. average?" The respondents could choose either "cheaper" or "more expensive." The second part of the question was: "How much [cheaper/more expensive] is the [Metro Name] metro area than the U.S. average?" Respondents could answer this second question with a drop-down menu ranging from 0% to 50%, in 1 percentage point increments.

We also provided an introduction for the question about earnings rank: "Now we want to

[5]When individuals were listing the second program, we required respondents to make a selection from a different metro area because otherwise no differences would be present in relative income and cost of living across choices. Our survey data indicates that no more than 4% of individuals tried to select the same metro area. For those subjects, the comparison was between two of their top programs but not necessarily the top two.

[6]Even though there are no large income differences in residency salaries, there can be large differences in post-residency salaries, especially across specialties.

ask you about your expected earnings rank. This rank is defined as the share of the working individuals of a city who earn less than you. You probably noticed that the distribution of earnings is different across different cities. As a result, with the same earnings, you may be relatively rich in some cities but relatively poor in other cities." After this introduction, we asked the following question for each city: "Imagine that you chose to work in [Metro Name]. With your individual annual earnings of $[Salary], you would be richer than what percentage of [Metro Name]'s individual earners?" Respondents could select their answer from a drop-down menu that ranges from "Richer than 1% of individual earners" to "Richer than 100% of individual earners," in 1% increments.

We focus on this definition of reference group because it is the most widely used approach in the related literature: e.g., Luttmer (2005) studies how the happiness of an individual is affected by the income of her neighbors.[7] In practice, individuals may care about their ranking in finer reference groups: e.g., they may care disproportionately about their relative standing with respect to neighbors in the same age cohort, rather than caring about all neighbors equally. However, this source of measurement error is not a major source of concern, to the extent that it can only introduce attenuation bias.

2.2.3 Information-Provision Experiment

One limitation with using perceptions is the potential for omitted-variable bias. For instance, conditional on income and perceptions about cost of living, perceptions about relative income may happen to be correlated with perceptions about other characteristics of the area, such as the crime rate, amenities, public goods, and so forth. To address this concern, we generate exogenous variation in the perceptions about cost of living and earnings rank by embedding an information-provision experiment in the survey.

Immediately after respondents provided their prior beliefs on both measures, they were shown two messages: one page with statistics about the cost of living in the two cities being considered and a second page with statistics about the earnings rank in each of the two cities. The following message is a sample of the feedback page for cost of living: "Los Angeles-Long Beach-Anaheim, CA metro area is 17.0% more expensive than the U.S. average. The Champaign-Urbana, IL metro area is 6.6% cheaper than the U.S. average." The following message is a sample of the feedback page for earnings rank: "With your individual annual earnings of $54,000, you would be richer than 57.9% of Los Angeles-Long Beach-Anaheim, CA's population. With your individual annual earnings of $54,000, you would be richer than

[7] Moreover, this geographic definition of references group is used more generally in the literature of social interactions more generally: e.g., Perez-Truglia (2017) and Perez-Truglia and Cruces (2017) study how an individual's political participation is affected by the participation of her neighbors.

60.3% of Champaign-Urbana, IL's population." In both of these feedback pages, individuals were asked to take a moment to review the information carefully and were alerted that the information was only going to be shown once. We did not allow respondents to continue to the next page until at least 10 seconds had elapsed.[8]

After individuals finished reviewing the feedback, we re-elicited their perceptions about cost of living and earnings rank, which we denote as the posterior beliefs. Given that our feedback entailed many figures for participants to remember and process, we wanted to make it easier for individuals to compare the options. Therefore, after eliciting respondent's posterior beliefs, we gave subjects a third page of feedback based on their posterior beliefs. The following is a sample of that feedback page: "We understand this is a lot of information to process, so we will help you make the comparison simpler. According to your final answers about incomes, cost of living and earnings rank: If you chose to live in Los Angeles-Long Beach-Anaheim, CA, you would be able to afford 19.7% less than if you chose to live in Champaign-Urbana, IL. If you chose to live in Los Angeles-Long Beach-Anaheim, CA, your earnings rank would be 3.3% lower than if you chose to live in Champaign-Urbana, IL."[9]

We computed the statistics shown to the subjects using two alternative data sources, and we cross-randomized which of the two sources were shown to each individual. The sources were randomized between individuals; that is, we used the same cost of living source for the two cities being considered by each individual, and the same earnings data source for the two cities. As a result, individuals were randomly assigned to one of four treatment groups. For cost of living estimates, the two sources used were the Regional Price Parity (RPP) data by the Bureau of Economic Analysis and the Cost of Living Index (COLI) data compiled by the Council for Community and Economic Research. For the earnings rank feedback, the two sources used were the American Community Survey (ACS) and the Current Population Survey (CPS), both conducted by the U.S. Census Bureau.[10]

This source randomization created a substantial amount of exogenous variation in signals. For instance, the correlation of the pairwise difference in cost of living shown to the respondents versus the corresponding pairwise difference from the alternative source is 0.656; the corresponding correlation for the earnings rank is 0.649. These differences across sources arise from a combination of several factors, most notably sampling variation and data definitions.

[8]The median time spent on the feedback page was 18.5 seconds.

[9]The difference in cost of living was calculated as $100 \cdot \left(\frac{w_1}{w_2}\frac{COL_2}{COL_1} - 1\right)$, where w_i is the nominal wage for city i and COL_i is their posterior belief about cost of living (from 50 to 150). The difference in earnings rank was calculated as $100 \cdot \left(\frac{ER_1}{ER_2} - 1\right)$, where ER_i is the posterior belief about earnings rank in city i. As with the other feedback pages, 10 seconds had to elapse before respondents could move to the next page. The median duration on the post feedback page was 19.5 seconds.

[10]For more details, see Appendix C.

For instance, the cost of living data is subject to sampling variation because it tracks the prices of a limited number of goods and services, and earnings rank data is subject to sampling variation because the estimates are based on a limited number of survey respondents. The variation in definitions arise because different cost of living indices give different weights to expenditure categories, and because the earnings rank measures are based on surveys with significant differences in the survey method and the phrasing of the questions used to elicit total annual earnings.

For the sake of transparency and to ensure the validity of the information, the individuals were debriefed in the feedback messages on the name of the source of the information that they received. We would not expect the source name to have an effect in and of itself, given that the individuals did not have expertise on the data, and even experts may have only a weak preference on which source is more trustworthy depending on the application. Indeed, we find that the reaction of individuals to the information was orthogonal to the name of the information source.[11]

2.2.4 Rank Submission Choices

The survey asked respondents to indicate which program they expected to rank higher when submitting to the NRMP: "As of this moment: of the two programs discussed so far, which one would you expect to rank higher for the NRMP?" Individuals could indicate their ranking on a 6-point scale ranging from "Very likely [Program 1] (in [Metro 1])" on one side to "Very likely [Program 2] (in [Metro 2])" on the other. In the baseline results we look at the binary choice of whether they expect to rank Program 1 over Program 2 because a comparison with the ex post submission choices is more straightforward. Nevertheless, results are similar when using the full likelihood scale.[12]

The algorithm used by NRMP was designed by Roth and Peranson (1999) to be 100% resistant to attempts of "strategic behavior," meaning that it is a weakly dominant strategy for students to submit their true preferences (i.e., it is optimal regardless of the behavior of the other applicants).Students receive training from the NRMP that makes it explicit that it is in their best interest to submit truthful ranks. Indeed, survey data indicates that only 5% of participants attempt to misreport their true preferences with a strategic motive (Benjamin et al. 2014; Rees-Jones 2017).[13] Furthermore, almost all NRMP participants

[11]Results reported in Appendix Figure D.3.

[12]Results reported in Table D.6.

[13]These results are consistent with other surveys (NRMP 2015). Given the small share of individuals attempting to manipulate rankings, we decided not to include questions about this. Relatedly, Rees-Jones and Skowronek (2017) provide complementary behavioral evidence that NRMP participants may fail to fully optimize, including a discussion of the source of those frictions.

receive a match,[14] and backing out from a match entails serious sanctions.[15] As a result, the rank choices provide a direct proxy for the individuals' true preferences.

Since most of the evidence on relative concerns is based on the happiness literature (e.g., Luttmer 2005), we want to compare preferences inferred from choice data with respect to the preferences inferred from happiness data, in the spirit of Benjamin et al. (2012; 2014). For this purpose, we included the following question about happiness rank: "If assigned to it, in which of the two programs would you expect to live a happier life?" Responses used the same likelihood scale as for rank.

2.3 Follow-Up Survey

Shortly after the NRMP rank submission window closed, we conducted a follow-up survey with the subjects that responded to the baseline survey. Appendix A.2 shows the full questionnaire of the follow-up survey.

Most importantly, at the very beginning of the survey we collected data on the final rank order submitted to the NRMP. Additionally, we took the opportunity to ask individuals for some additional information. We elicited the perceptions about cost of living and earnings rank, which allows us to measure the persistence of the information learned in the information-provision experiment. Also, we measured additional characteristics of the subjects, such as the places where they grew up and measures of materialism (Richins and Dawson 1992) and competitiveness (Smither and Houston 1992). We did not measure these secondary characteristics in the baseline survey due to space and time constraints.

3 Econometric Model

3.1 Baseline Model

In this baseline model, we exploit all the variation in perceptions of earnings rank and cost of living, which includes the experimental variation induced by our information-provision as well as the remaining non-experimental variation.

Let i index subjects and $j \in \{1, 2\}$ denote the two programs being considered by the subject. We define $ER_j^{i,posterior}$ and $COL_j^{i,posterior}$ as the posterior beliefs for earnings rank

[14] For instance, 95% of the 27,048 U.S. graduating medical students received a successful match in 2017.

[15] For example, applicants with confirmed violations of NRMP policies are subject to a one year bar from accepting or starting a position in any program sponsored by a Match-participating institution, from one year to a lifetime bar from participation in future NRMP Matches, and from one year to a lifetime identification in the matching system as a match violator (Source: http://www.nrmp.org/policies/the-match-commitment/). Additionally, the NRMP has established rules prohibiting programs from contacting candidates to ask or coordinate their rank orders.

and cost of living for program j in the baseline survey. Let $ER_{1,2}^{i,posterior} = ER_1^{i,posterior} - ER_2^{i,posterior}$ be the perceived difference in earnings rank between the two programs. Similarly, let $COL_{1,2}^{i,posterior} = COL_1^{i,posterior} - COL_2^{i,posterior}$ be the perceived difference in cost of living between the two programs. Let $Program_1 \succ_i \; Program_2$ denote that individual ranks program 1 over program 2, and let $I(\cdot)$ be an indicator function. The regression specification is:

$$I(Program_1 \succ_i \; Program_2) = I\Big(\beta^{ER} \cdot ER_{1,2}^{i,posterior} + \beta^{COL} \cdot COL_{1,2}^{i,posterior} + \\ + \theta X^i + \varepsilon_i \geq 0\Big), \quad (1)$$

where X^i is a vector of control variables and θ is the corresponding vector of coefficients. We always include a constant and the log-difference of nominal residency wages as control variables. In the baseline specification, we include an additional set of controls consisting of pairwise differences in some residency and location characteristics: residency program rank (from Doximity), quality of life inferred from compensating differentials (Albouy 2016), population size, population density, share of African-American residents, share of Democrat residents, and share of urban population.[16] In any case, we present results with alternative sets of control variables.

In the baseline specification, we estimate a Probit model, which implies that the error term (ε_i) is normally distributed. As is typical in discrete-choice models, using a Probit model is convenient in the sense that the ratio between parameters can be readily interpreted as marginal rates of substitution. However, this specification choice is irrelevant in practice: the results are virtually identical if we use alternatives such as Logit or Linear Probability models.

The two key parameters of interest are β^{ER} and β^{COL}. The parameter β^{ER} measures preferences for relative income over the duration of the residency. Depending on the mechanism at play, we may expect β^{ER} to be positive or negative. For instance, the status models predict that $\beta^{ER} > 0$ (individuals want to choose less affluent ponds) while some social interaction models predict that $\beta^{ER} < 0$ (individuals want to choose more affluent ponds). The parameter β^{COL} measures preferences for purchasing power during the residency. We expect $\beta^{COL} < 0$: i.e., individuals prefer to live in places where they can afford to consume more.

Note that earnings rank and cost of living *after* the end of the residency would be part of the error term. The duration of a residency depends on the specialty: it lasts for a minimum of three years, it typically takes five years, and in some cases it may require a minimum of

[16] The source for the demographic characteristics is the 2011-2014 American Community Survey. For the share of Democrat residents, we use the share of Obama voters between all voters in the 2008 Presidential Elections.

seven years.[17]

In Section 8 we discuss the potential interpretations for β^{ER} and β^{COL}. One possible interpretation, which happens to be the one that motivated this survey design, states that β^{ER} and β^{COL} reflect preferences over relative and absolute consumption during the residency. Let absolute consumption be the nominal earnings divided by the cost of living index, and let relative consumption be the individual's rank in the distribution of absolute consumption in the same city. If the cost of living decreases in an area, it increases one's absolute consumption level because one can afford more goods with the same nominal income. However, it also increases the absolute consumption level of everyone else in the city, leaving one's relative consumption unchanged. In contrast, with the cost of living held constant, a change in the distribution of the earnings in a metro area affects one's relative consumption, but it does not affect one's absolute consumption. As a result, β^{ER} could be interpreted as the marginal utility from relative consumption, while $-\beta^{COL}$ could be interpreted as the marginal utility from absolute consumption. Furthermore, the ratio $-\frac{\beta^{ER}}{\beta^{COL}}$ would correspond to the marginal rate of substitution between relative consumption and absolute consumption.

3.2 Instrumental Variables Model

The second model exploits the variation in beliefs induced by the source-randomization experiment to estimate the causal effects of perceptions on choice. Let $ER_{1,2}^{i,shown}$ be the information randomly chosen to be shown to the individual, and $ER_{1,2}^{i,alt}$ be the alternative information that could have been shown to the individual, but was not shown. Let $\Delta ER_{1,2}^{i} = ER_{1,2}^{i,shown} - ER_{1,2}^{i,alt}$ be the difference between the information shown and the alternative information that could have been shown. We estimate an IV-Probit model that uses $\Delta ER_{1,2}^{i}$ and $\Delta COL_{1,2}^{i}$ as instrumental variables. In other words, this model uses the variation introduced by the random assignment of sources to estimate the effect of perceptions on choice:

$$\begin{aligned} I(Program_1 \succ_i Program_2) = I(&\beta^{ER} \cdot ER_{1,2}^{i,posterior} + \beta^{COL} \cdot COL_{1,2}^{i,posterior} \\ &+ \lambda_1 \cdot ER_{1,2}^{i,alt} + \lambda_2 \cdot COL_{1,2}^{i,alt} + \theta X^i + \varepsilon_i \geq 0) \end{aligned}$$

$$ER_{1,2}^{i,posterior} = \gamma_1^{ER} \cdot \Delta ER_{1,2}^{i} + \gamma_2^{ER} \cdot \Delta COL_{1,2}^{i} + \gamma_3^{ER} \cdot ER_{1,2}^{i,alt} + \gamma_4^{ER} \cdot COL_{2,1}^{i,alt} + \gamma_5^{ER} X^i + \epsilon_{1,i}$$

$$COL_{1,2}^{i,posterior} = \gamma_1^{COL} \cdot \Delta ER_{1,2}^{i} + \gamma_2^{COL} \cdot \Delta COL_{1,2}^{i} + \gamma_3^{COL} \cdot ER_{1,2}^{i,alt} + \gamma_4^{COL} \cdot COL_{2,1}^{i,alt} + \gamma_5^{COL} X^i + \epsilon_{2,i}$$

[17] A small minority of subjects may expect to continue living in the same city after the residency, in which case the cost of living and the distribution of earnings may also be relevant for the post-residency period.

There is a simple way to understand the intuition behind this instrumental variables approach. In a deceptive design, subjects would be shown the statistic from a certain source, but with random noise added to this statistic. Then we would only exploit the variation in beliefs generated by the random noise. In our context, $\Delta ER^i_{1,2}$ and $\Delta COL^i_{1,2}$ play the role of the random noise added to the feedback, only that they are generated in a non-deceptive manner.

4 Implementation Details and Summary Statistics

Our recruitment strategy is similar to that of Benjamin et al. (2014). During December 2016 we contacted the Associate Dean of Student Affairs at all 135 accredited medical schools in the United States by email to ask for permission to invite fourth year students participating in the 2017 Main Residency Match to take part in our study (a sample of the invitation email is shown in Appendix A.3). Our goal was to recruit as many respondents as possible, so we followed up, by email and phone, with all the deans who showed interest. Of the 79 schools that answered our invitation, 27 agreed to participate. The main reason given by the schools that declined to participate was school policy restricting external surveys, in place to avoid survey fatigue. Our sample of participating schools includes 22 of the 50 U.S. states, and it is quite representative of the whole sample of 135 accredited medical schools – we do not find statistically significant differences in observable characteristics such as total enrollment, average MCAT scores, undergraduate GPA at admission, acceptance rate, and U.S. News rank.[18]

For confidentiality reasons, we were not given email lists to directly invite students to participate in our study. Instead, the deans agreed to forward our invitation email containing the link to the survey to eligible students (i.e., senior medical students participating in the NRMP). This email invitation, a sample of which is shown in Appendix A.4, asked students to participate in a confidential survey about the Main Residency Match for a study on how medical students select residency programs. The message mentioned that the survey would take less than 10 minutes to complete and respondents would be sent a $10 Amazon gift card by email as a token of appreciation. Finally, the email stressed the eligibility criteria for participating in the survey: being a graduating medical student participating in the Main Residency Match who has not yet submitted his or her rank to the NRMP.[19]

The only reason why we excluded individuals who had previously submitted their ranks

[18] For details, see Appendix B.

[19] There are a number of alternative matches for some specialties that have different deadlines than the Main Residency Match.

was because we wanted individuals who were still deciding and thus prone to using the signals from the information-provision experiment. However, this concern is not important in the sense that submissions can be modified anytime before February 22. Even if some students had already submitted their rank at the time of responding to the survey, they would still be able modify their rank. In any case, the vast majority of our subjects responded to the baseline survey quite early in the submission period.

We took several measures to minimize the chance that non eligible students would participate in the survey. First, deans were asked to carefully forward the invitation to senior students participating in the Main Residency Match. This request was not an issue since such a mailing lists already existed; targeted announcements were already being sent to this group during the semester regarding the Match. Second, individuals were reminded of these restrictions in the invitation email and on the consent page of the survey. Third, the first questions of the surveys acted as filters; we asked what match the respondent was participating in and whether they had already submitted their ranks. If they responded with a match other than the Main Residency or "yes" to already submitting their rank, the survey ended there, and they were excluded from taking the survey again.[20]

Last, at the end of the survey, respondents were required to submit their university email address to "sign" a statement claiming that they were medical students participating in the NRMP and they understood that we reserved the right to verify their status before making a payment. We were able to confirm the validity of 100% of respondents for a subset of schools. Given all the measures taken and the evidence obtained, we are confident that the survey data are of high quality.

The invitation emails were sent to students in a staggered way, with the first round of invitations sent on January 6, 2017, and the last round of invitations and reminders sent on February 7, 2017. We estimated that the student invitations were forwarded to around 3,676 students in total, with 1,080 finishing the baseline survey, implying an overall response rate of 29.38%. The median survey completion time was almost 9 minutes. At the end of the baseline survey we included an attention check question that was passed by 96.4% of respondents. For the sake of transparency, we do not drop the group that did not pass the attention check – indeed, we do not drop any other group from the baseline sample.[21]

On February 23, 2017, the day after the NRMP rank submission deadline, we sent respondents who participated in our baseline survey an invitation to participate in the follow-up

[20]The survey platform blocks users from taking the survey again by using their I.P. address and cookies, although students could circumvent this restriction by opening the survey link from a different device.

[21]Upon inspection of the data, the 3.6% of respondents who failed the attention check seem to have answered the survey as consistently as everyone else. And, as reported in Appendix Table D.5, the results are virtually the same if we drop this 3.6% of the sample.

survey. We offered participants an additional $5 Amazon gift card for participating in this shorter follow-up survey. We closed the follow-up survey on March 12, one day before Match Week started (i.e., the time when the students find out where they are matched). The response rate to the follow-up survey was 90.62%. Moreover, the characteristics of the individuals who responded to the follow-up survey are similar to the characteristics of individuals who did not respond to the follow-up.[22]

Figure 1 presents the distribution of dates when subjects responded to the baseline survey, when they responded to the follow-up survey, and when they submitted their ranks to the NRMP (for those who provided this information in the follow-up survey). On average, students responded to the baseline survey 24.5 days (s.d. 12.9) before submitting their ranks, and responded to the follow-up survey 13.9 days (s.d. 11.8) after submitting their ranks.

Figure 2 shows the geographic distribution of the metropolitan areas in which students' top-two programs are located. This figure shows that there is a broad geographical coverage of the U.S. territory.

Table 1 provides descriptive statistics for the key variables used in the analysis. Column (1) corresponds to all respondents to the baseline survey. Around 48% of respondents were male, the average age was 27 years, 35.4% of respondents were single, 23.9% were married, and 40.7% were in a long-term relationship. On average, students were offered a salary of $54,000 for the first year of their residency – this salary would make them richer than 56% of earners in the average metro area. Of course, this sample is not representative of the general U.S. population of adults: most notably, our subject pool is younger and more educated. Nevertheless, our subject pool is close to the U.S. average in terms of nominal wages and gender composition.[23]

To verify that the randomization was successful, Table 1 breaks down the descriptive statistics by each treatment group. This table also reports the p-value for the test of the alternative hypothesis that at least one mean is different across the four treatment groups. First, this table shows that the number of respondents was almost identical number of respondents across all groups. Second, this table shows that the differences in individual characteristics are economically small and statistically insignificant across the treatment groups, thus confirming that the random assignment was successful.

[22]Results presented in Appendix Table B.3.

[23]For more details, see Appendix Table B.4.

5 Results: Distribution of Perceptions and Learning

5.1 Variation in Nominal Income, Cost of Living and Earnings Rank

We first show that there is enough variation in cost of living and earnings rank to allow for the estimation of the key parameters. Given that we do not observe the "true" cost of living or earnings rank, but imperfect estimates based on different data sources, the following results use our "baseline" estimates: the RPP measure of cost of living and the ACS measure of earnings rank (using the alternative data sources yield similar results).

Figure 3 shows a scatterplot of the pairwise differences in cost of living vs. the differences in earnings rank. This figure shows three facts that are crucial for estimating the preferences for cost of living and relative income. First, the substantial dispersion in the y-axis suggests that there are large differences in cost of living across the pairs of cities that the individuals must choose from. Second, the substantial dispersion in the x-axis suggests that there are large differences in earnings rank across the pairs of cities that the individuals must choose from.[24] Third, the $R^2 = 0.22$ indicates that, even though the two are correlated,[25] substantial orthogonal variation exists between cost of living and relative income.

5.2 Distribution of Prior Beliefs

To the best of our knowledge, ours is the first paper to measure perceptions about cost of living and earnings ranks across different cities. To get a sense of how informed individuals are about these aspects of their decision-making, we start by comparing their prior beliefs (that is, perceptions prior to the feedback) to the baseline statistics: the RPP measure of cost of living and the ACS measure of earnings rank.

Respondents seem to have a relatively good idea of the cost of living in the cities they are considering. Figure 4.a shows respondents' prior beliefs about cost of living along with the corresponding RPP estimates. The RPP is meant to reflect all sources of expenditures, and for that they employ data on prices in: apparel, education, food, housing, medical, recreation, rents, transportation and other goods and services. If answers were completely accurate, we would expect to see all responses on the 45 degree line. On average, prior beliefs overestimate the baseline estimate by just 4 percentage points; and the prior belief and RPP estimates are positively correlated, with an R^2 of 0.550.

[24] Furthermore, the vast majority of these differences in cost of living and earnings rank are orthogonal to differences in nominal income – see Appendix D.1 for details.

[25] The slope of -0.664 suggests that, on average, relatively more expensive cities tend to have a higher distribution of nominal earnings.

However, individuals are substantially less well informed about their earnings ranks. Figure 4.b plots prior beliefs about earnings rank against the ACS estimates. On average, individuals underestimate earnings ranks by almost 16 percentage points; and the prior belief and ACS estimates are positively correlated, but with an R^2 of just 0.029. Because we are ultimately interested in relative differences for their decision making, we repeat this exercise using pairwise differences instead of levels. It seems that respondents have a better understanding of relative differences in earnings rank, though they still remain far less accurate than perceptions over cost of living.[26] This finding suggests that, while prior evidence suggests that individuals have significant biases when assessing their position in the national income distribution (Cruces et al. 2013; Karadja et al. 2017), these biases are even more substantial when individuals try to predict their position in places where they are not currently living.

5.3 Learning from Statistics

We next examine whether respondents learned from the information we provided. To do this, we examine the relationship between the initial perception gap of respondents (i.e., the signal received minus the prior belief) and the extent to which they revise their responses (the posterior belief minus the prior belief). If respondents learn from the information provided, we would expect a positive relation between their perception gaps and their revisions; that is, respondents who originally overestimated would revise their beliefs downwards, while those who underestimated would revise in the opposite direction.

Indeed, the slope between the perception gaps and revisions can be used to quantify the degree of learning from information.[27] Let b_k^{prior} denote the mean of the prior belief k, b_k^{signal} the signal about k, and $b_k^{posterior}$ the mean of the corresponding posterior belief. When priors and signals are normally distributed, Bayesian learning implies that the mean of the posterior belief should be a weighted average between the signal and the mean of the prior belief:

$$b_k^{posterior} = \alpha_k \cdot b_k^{signal} + (1 - \alpha_k) \cdot b_k^{prior}$$

The degree of learning can be summarized by the weight parameter α_k. This parameter can take values from 0 (individuals ignore the signal) to 1 (individuals fully adjust to the signal). We can rearrange the previous equation:

[26] Detailed results reported in Appendix D.2.

[27] For a discussion about the estimation of learning models with survey experiments, see Armantier et al. (2016) and Cavallo et al. (2017). Also, in relation to the identification of preferences from information-provision experiments, see Wiswall and Zafar (2014).

$$b_k^{posterior} - b_k^{prior} = \alpha_k \cdot \left(b_k^{signal} - b_k^{prior}\right)$$

Which implies that we can estimate the learning rate (α_k) by estimating a regression of the revision ($b_k^{posterior} - b_k^{prior}$) on the perception gap between the prior and the signal ($b_k^{signal} - b_k^{prior}$).

Respondents strongly updated their beliefs after being provided with feedback. Figure 5 presents the reduced-form effects of information for cost of living and earnings rank, respectively. Figures 5.a and 5.d present the short-term effect, that is, the revision made by respondents directly after being given the information. The short-term learning rates, given by the slopes reported in these figures, are 0.879 (s.e. 0.010) for the cost of living and 0.873 (s.e. 0.011) for the earnings rank. These two learning rates are statistically significant, precisely estimated, and we cannot reject the null hypothesis that they are equal to each other (p-value=0.754). These learning rates are remarkably close to 1, meaning that respondents almost fully reacted to the signals.

One limitation with this evidence is that individuals may have revised their beliefs towards the truth regardless of the feedback we provided. For instance, they may have taken extra time to think about the question, leading to a more accurate response. The source experiment was designed to test this specific hypothesis. We construct two variables: the information actually shown and the "alternative" information that could have been shown. If the alternative information had any effect beyond the information shown, that would be evidence that part of the revisions were due to reversion to the truth rather than reversion to the information provided. Figures 5.b and 5.e show the relation between the alternative information and the revision adjusted for the information actually shown. The alternative information indeed has no effect: the coefficients are close to zero (-0.034 for cost of living and 0.060 for earnings ranking) and precisely estimated. Furthermore, in Section 8.1 we show that there was no cross-learning (i.e., feedback on cost of living did not affect beliefs about earnings ranking and vice versa).

In survey experiments, one main concern is that instead of inducing genuine learning, the information provided in the experiment may elicit spurious reactions. For instance, if an individual is told that the cost of living in a city is "10% more expensive than the U.S. average" and then later asked about the cost of living in the same city, he or she may report a cost of living that is closer to "10% more expensive than the U.S. average" for spurious reasons, such as unconscious numerical anchoring (Kahneman and Tversky 1972). Under the assumption that these effects are temporary, we can disentangle genuine from spurious learning by looking at the reaction to the information provided in the experiment that persisted over time (Cavallo et al. 2017).

We look at the persistence of the effect of feedback between the time participants responded to the baseline and follow-up surveys, which was 38.4 days on average. Figures 5.c and 5.f show the relation between the initial perception gap and the long-term revision based on beliefs reported in the follow-up survey (i.e., $b_k^{posterior,LT} - b_k^{prior}$). There is substantial persistence the effects of the feedback: the estimated slope for the initial perception gap and the long-term revision (i.e., the difference between long-term belief and the initial prior belief) for cost of living is 0.752 (s.e. 0.016), while for earnings rank it is 0.626 (s.e. 0.020). These longer-term revisions are slightly weaker than the short-term revisions, but that result is expected given that individuals must have gathered some additional information in the time between the two surveys.

6 Results: Preferences for Relative Income

6.1 Average Preferences

We first explore the baseline estimates of the effects of earnings ranking and cost of living. The baseline specification uses the Probit model from Section 3, with the expected rank submission as dependent variable. This specification exploits all the variation in perceptions, which includes the experimental variation induced by our information-provision as well as the remaining non-experimental variation. We introduce the experimental estimates later in this section.

Respondents prefer a lower cost of living. The Probit coefficients are presented in Table 2. Column (1) presents the results for the full sample, while columns (2) through (7) present results by demographic subgroups. The estimated β^{COL} from column (1) is negative and statistically significant (p-value=0.027), suggesting that the average individual prefers programs with lower costs of living. To better understand the magnitude of these Probit coefficients, we can transform them into the corresponding marginal effects, where increasing the cost of living by 1 percentage point at a program's location decreases the probability of choosing that program by 0.201 percentage points (which can be interpreted as a behavioral elasticity of -0.201).[28]

The fact that medical students care about cost of living during the residency is consistent with the view that money is a primary motivation for doctors. For instance, according to a 2008 survey, 49% of pre-med students self-reported being primarily motivated by money in their career choice (Daniel and O'Brien 2008).[29] Even though β^{COL} is statistically and

[28]These marginal effects are reported in Appendix Table D.1.

[29]These survey results are based on responses from 461 takers of the Kaplan MCAT test in February 2008 and 453 takers of the Kaplan LSAT test in February 2008.

economically significant, it does not imply that cost of living during residency is the main feature that medical students pay attention to. Intuitively, if doctors care about their post-residency consumption, they should choose residencies that offer better post-residency job prospects. Indeed, in complementary analysis, we find that doctors care substantially more about the prestige and career prospects than about the cost of living during their residency.[30]

Most important, the average subject also prefers a higher earnings rank: the estimated β^{ER} from column (1) is positive and statistically significant (p-value=0.065). This coefficient suggests that the average individual prefers to live in a city where, holding her cost of living constant, she earns more than her neighbors. The corresponding marginal effect indicates that increasing the earnings rank at a program's location by 1 percentage point increases the probability of choosing that program by 0.186 percentage points (for a behavioral elasticity of 0.186). The elasticity for cost of living (-0.201) is similar in magnitude to the elasticity for earnings rank (0.186) – indeed, their difference is statistically insignificant. This finding suggests that individuals care about relative income nearly as much as they care about cost of living.

The evidence suggests that individuals take relative income into account when they make their location decisions. This preference could be the product of a combination of multiple mechanisms, pushing in different directions. The positive sign of β^{ER} suggests that the dominant mechanism is consistent with models in which richer neighbors impose a negative externality, as in Boskin and Sheshinski (1978), Frank (1985), Cole et al. (1998) and Luttmer (2005), among others. In this section, we focus on identifying the preference parameters, and we provide a discussion of the interpretation of these parameters in Section 7 below.

6.2 Heterogeneity by Relationship Status

The average preferences could potentially mask substantial heterogeneity. For instance, Luttmer (2005) finds that the effect of relative income on happiness is driven entirely by the sample of non-single individuals. Furthermore, evidence from the urban economics literature indicates that single and non-single individuals have different locational preferences (e.g., Couture and Handbury 2016; Gautier et al. 2010). To explore heterogeneity in preferences, columns (2) through (7) of Table 2 present estimates broken down by the basic demographic groups measured in the baseline survey: relationship status, gender and expected (post-residency) income.

To explore heterogeneity by relationship status, we elicited the relationship status using the same categories as in Luttmer (2005). Column (2) of Table 2 shows the effect for non-

[30]Results presented in Appendix D.7.

single individuals (i.e., the 65% of the sample who are married or in a long-term relationship) and column (3) for the sample of single individuals (35% of the sample).[31] It is important to note that by non-single we only refer to their relationship status, not to whether the respondent participates as a dual match, which is a special regime used by roughly 7% of subjects—indeed, the results are similar if we drop subjects with dual matches.[32]

Comparing columns (2) and (3) indicate large heterogeneity in β^{ER} by relationship status. For non-single individuals, the estimated β^{ER} (2.236) is positive and statistically significant at the 1% level. For the sample of single individuals, β^{ER} (-1.538) is negative and statistically significant at the 10% level. The direction of the difference in relative concerns between non-singles and singles is consistent with the evidence from Luttmer (2005).

The difference in β^{ER} between non-singles and singles is highly statistically significant (p-value=0.001). Moreover, to address spurious results from multiple hypothesis testing, for each p-value reported in the table we also report the corresponding q-value based on Benjamini and Yekutieli (2001). The q-value indicates the minimum false discovery rate (i.e., the expected proportion of rejected null hypotheses that are actually true) at which the null hypothesis would be rejected for that test given all tests reported in the same table. The difference in β^{ER} between singles and non-singles has a q-value of 0.030, which indicates that this heterogeneity is unlikely to be spurious. Contrary to the case of preferences for earnings rank, the relationship status does not seem to affect the preferences for cost of living. According to columns (2) and (3) of Table 2, the estimated β^{COL} is -1.087 for non-singles and -1.058 for singles, with the difference being statistically insignificant (p-value=0.977).

These estimates suggest that while non-single individuals prefer to live in less affluent ponds, single individuals would rather live in more affluent ponds. While the preferences of non-single individuals can be rationalized by status models, the preferences of single individuals cannot be rationalized by such models. One potential explanation for the preferences of single individuals lies in their local social interactions, such as in the dating market. These subjects are at their prime dating age, and thus are likely to be looking for long-term partners during their residency. Since individuals prefer to date rich partners (Fisman et al. 2006, Hitsch et al. 2010), this can naturally create a preference for locating in more affluent ponds. Moreover, prior evidence suggests that, relative to single men, single women may have a stronger preference for finding rich partners (Bertrand et al. 2015; Bursztyn et al. 2017b). Consistent with this view, we find that the preference for more affluent ponds among singles is driven primarily by single women, although this result is imprecisely estimated.[33]

[31] Appendix Table D.3 shows results breaking down the non-single individuals into married and in a long-term relationship. The relative concerns are similar between these two groups.

[32] See Appendix Table D.5 for more details.

[33] Results reported in Appendix D.3.

Relative to other random single individuals from the general U.S. population with a similar salary, the singles in our subject pool may have a stronger reason to seek more affluent ponds: even though their wages during the residency put them near the middle of the U.S. distribution of earnings, their expected *post-residency* earnings will place them near the top of the earnings distribution. If these subjects have a desire to meet a partner that can match their permanent income, they should try to locate in the richest areas of the United States.

Beyond dating preferences, there may be gender differences in preferences for cost of living and relative income. For instance, there may be gender differences in consumption aspirations or in status concerns. Columns (4) and (5) explore potential differences in preferences by gender. These gender differences are small: β^{ER} is similar for females (1.041) and males (0.896), and β^{COL} is also similar for females (-0.972) and males (-1.443). Moreover, neither of these two differences are statistically significant (p-values of 0.894 and 0.642, respectively).

Last, even though all these subjects receive a similar income during the residency, they have very different expected incomes after they finish their residencies. It is possible that individuals who selected high-earning specialties may be more concerned about relative income. The test this hypothesis, columns (6) and (7) split the sample in specialties with above and below median post-residency average salaries. Again, the differences in coefficients are statistically insignificant: β^{ER} is 1.433 for below-median specialties and 0.777 for above-median specialties, and β^{COL} is -0.690 for below-median specialties and -1.238 for above-median specialties, with neither of those differences being statistically significant (p-values of 0.544 and 0.580, respectively).

We also computed heterogeneity by other characteristics measured in the follow-up survey. None of these dimensions are nearly as important as relationship status for predicting heterogeneity in preferences for income, both in terms of magnitude and statistical significance.[34] Because of the magnitude of the heterogeneity by relationship status, in the remainder of the paper, we report estimates for non-single and single respondents, in addition to estimates for the entire sample.

6.3 Robustness Checks: Controlling for Other Observable Characteristics

One potential concern with the baseline specification is that of omitted-variable biases. For instance, if places where an individual expects higher earnings rank (i.e., less affluent metro areas) are systematically worse in terms of quality of life, then failing to account for quality

[34] Results reported in Appendix Table D.4.

of life would introduce a negative bias in β^{ER}, thus making relative concerns look weaker than they actually are.

We present the baseline estimates using alternative sets of control variables in Table 3. Each row corresponds to a different regression, with a different set of control variables. The first row presents results for our baseline specification, but without including any control variables for the characteristics of the program or the metro area. The second row corresponds to the baseline specification from Table 2, which includes the six baseline controls listed in Section 3.1. The results in the first two rows of Table 3 indicate that β^{ER} and β^{COL} are qualitatively and quantitatively similar between the baseline specification and the specification without controls.

The third through last rows of Table 3 include different sets of additional controls. These sets of controls were selected based on attributes that could potentially be relevant for the options of the subjects and at the same time may be correlated to the earnings rank. For instance, we may want to control for crime rates: living in a less affluent city may be desirable for medical students interested in certain specialties where they must learn to treat injuries that are more common in high-crime areas, such as gunshot wounds. Also, we may want to account for place of origin: medical students, who tend to grow up in affluent areas, may want to remain in the the same areas where they grew up (Agarwal 2015).

We examine the following groups of attributes: demographic characteristics (population, population density, percentage urban population, percentage same gender, percentage age 25 to 34, share of college graduates, share foreign, share Hispanic, and share black); amenities (quality of life from Albouy 2016, per capita spending on local public goods, per capita spending on education and health, overall crime rate and violent crime rate, share of registered Democrat voters in the 2012 election); geography (distance of program to city where they grew up, distance of program to current medical school); economic factors (estimated income taxes, federal and state income taxes, local sales tax, rent prices, and the Gini coefficient); a set of state dummies; objective program characteristics (residency program rank from Doximity, dummies for university hospitals and for community hospitals), and subjective program characteristics (the subjective rank in prestige, purpose, and prospect, as reported in the follow-up survey).

Comparing the results across rows of β^{ER} and β^{COL} of Table 3 suggests that these estimates are robust to the choice of control variables, both in terms of statistical significance and economic significance. Of course, small differences occur in the point estimates across specifications. For instance, relative to the baseline β^{ER} of 0.995 for the entire sample in column (3), β^{ER} ranges from a minimum of 0.703 with all controls to a maximum of 1.199

with subjective program characteristics.[35] However, all of these differences are statistically insignificant.

6.4 Robustness Checks: Experimental Estimates and Long-Term Effects

In this section, we present results from two robustness checks. The first check addresses concerns about omitted-variable bias by exploiting the exogenous variation in beliefs generated by the source-randomization experiment. The second is intended to address potential concerns about spurious effects of the information-provision experiment, by comparing the short-term effects to the long-term effects. To make these estimates directly comparable to the long-term effects, in this section we restrict the sample to individuals who responded to the follow-up survey.

Panel A of Table 4 presents the results for β^{ER}. The first row presents the baseline specification, while the second row presents the experimental estimates. The experimental estimates are less precisely estimated than the baseline estimates because they only use a portion of the variation in beliefs. For each of the subgroups of single and non-single respondents, shown in columns (1) and (2), the estimated β^{ER} is qualitatively consistent across the baseline and experimental specifications. For non-singles, the coefficient is 2.380 (p-value=0.001) in the baseline specification vs. 2.977 (p-value=0.025) in the experimental specification. And for singles, the coefficients are -1.656 (p-value=0.095) in the baseline specification vs. -4.964 (p-value=0.012) in the experimental specification.

Column (3) shows that, for the entire sample, β^{ER} is slightly lower in the second row (0.867) than in the first row (1.141) and, due to the lower precision, becomes statistically insignificant in the second row. However, we must take this finding with a grain of salt. First, due to the precision of the experimental coefficient, this difference between the first and second rows is statistically insignificant. Second, the reduction in the average β^{ER} is driven primarily by the fact that the coefficient becomes more negative for singles. This highlights the importance of taking potential heterogeneity of preferences into account when studying location preferences.

Panel B of Table 4 presents the results for β^{COL}. The results from the baseline specification (first row) are qualitatively different from the results in the experimental specification

[35] According to the pseudo-R^2 reported in panel C of Table 3, including these variables increases the explanatory power of our model to some degree. For the full sample in column (9), the pseudo-R^2 increases from 0.015 in the specification with no additional controls to a minimum of 0.018 with controls for objective program characteristics or amenities, and a maximum of 0.123 with controls for subjective programs characteristics (or 0.218 with all controls).

(second row). All the coefficients (for the entire sample, singles and non-singles) become positive, are imprecisely estimated, and are statistically insignificant. We must take this evidence with a grain of salt: since the experimental estimates are not precisely estimated, we cannot rule out large negative values for β^{COL}, and in most cases we cannot reject that the experimental coefficients are equal to those from the baseline specification. Also, the coefficients from the first and second row are not be expected to be equal, to the extent that the experimental coefficients identify local average preferences instead of average preferences.[36] However, these findings are at least suggestive evidence that the baseline estimates may overestimate the importance of cost of living.

As discussed above, the treatment groups were balanced in observable characteristics, suggesting that the randomization was indeed successful. As an additional robustness check, we re-estimate the instrumental variables model but, instead of the rank order, we use the list order as dependent variable (i.e., the order in which the individual listed the residency programs at the beginning of the survey). Because it takes place before the provision of feedback, the feedback should not have any effect on the list order. We present results for this falsification in the fourth row of Table 4. As expected, the estimated values of β^{ER} and β^{COL} are close to zero and statistically insignificant, in the full sample as well as in the sub-samples of non-singles and singles.

The Appendix presents some additional results. In all the instrumental variable specifications, we strongly reject the null hypothesis of weak instruments. Also, the learning rates implied by the first-stage coefficients are always close to 1, and for that reason the instrumental variables estimates are similar to the reduced form estimates.[37]

The second robustness test is intended to address potential concerns about spurious effects of the information-provision experiment such as salience and experimenter-demand effects. For example, by asking individuals questions about the cost of living and earnings rank, the baseline survey makes those aspects more salient, which may make individuals overweight them in their expected choice. However, this salience effect may not necessarily exaggerate the importance of relative income, because they would be expected to inflate both β^{ER} and β^{COL}.

We should also be concerned about potential experimenter-demand effects: by providing individuals with information about cost of living and earnings rank, the experimenter may be putting pressure on the subjects to use this information in their expected choice. Again, this source of bias would not necessarily exaggerate the importance of relative income: since

[36] For instance, it is plausible that the information-provision experiment disproportionally affected individuals who were the most unsure about their priors beliefs about cost of living, who likely are those who care the least about cost of living.

[37] Reduced-form and first-stage estimates are presented in Appendix Table D.7.

most individuals do not want to reveal to others that they care about status (Shigeoka and Yamada 2016), the experimenter-demand bias would probably shrink β^{ER} towards zero.[38]

To address these remaining concerns, we estimate the effects of the information provision on the final rank submission, which takes place an average of 38.4 days after the information provision. This can be achieved by using the same instrumental variable model, but using the final submission rank (elicited in the follow-up survey) instead of the expected submission rank (elicited in the baseline survey) as the dependent variable. If the effects were spurious due to salience or experimenter-demand effect, we would expect that the information provided in the experiment would not have any effect on the final submission choice. In other words, we test whether the information provided a month before the submission deadline had a long-lasting effect on the final ranks submitted.

The third row of Table 4 presents the experimental estimates based on the long-term effects of the experiment. By comparing the coefficients in the third row to those from the second row, we can compare the short-term and long-term experimental effects. Panel A of Table 4 presents the results for β^{ER}, while panel B corresponds to β^{COL}. The long-term experimental coefficients are somewhat different from the short-term experimental coefficients, but those differences are mostly statistically insignificant. Most important, the coefficient on β^{ER} is still positive (1.993) and statistically significant for non-singles, and negative (-5.285) and statistically significant for singles.

7 Auxiliary Experiment

In the previous section, we present evidence that individuals consider relative income in their location choices. In this section, we present complementary evidence from an auxiliary experiment.

7.1 Survey Design

We designed a variation of the survey instrument, attached in Appendix D.15, that can be used in other contexts besides the medical residency match. At the beginning of this survey, we ask respondents to list two cities that they know well to which they would consider moving to. The rest of the survey instrument is identical to the baseline survey instrument from our main experiment with medical students: we elicit prior and posterior beliefs about cost of

[38]Also, our survey was conducted confidentially and online, which reduces the scope for experimenter-demand effects (Van Gelder et al. 2010). Additionally, it would be difficult to reconcile the experimenter-demand channel with the finding that the earnings rank had a positive effect on non-singles and a negative effect on singles.

living and earnings rank, we conduct the information-provision experiment, and we elicit preferences for the two cities under consideration.

We conduct this auxiliary experiment using a sample of respondents from Amazon Mechanical Turk.[39] Compared to the main residency match experiment, the primary difference is that the subjects in the auxiliary experiment are not moving anytime soon, so we cannot followup with them to measure the effects of the information provision experiment on their actual location choices. Instead, we measure the effects on their expected location choices. On the one hand, this is a limitation of the auxiliary experiment.[40] On the other hand, this auxiliary sample has the advantage that, due to a wide availability of subjects, it is possible to run additional experiments at any time.

The results from this additional experiment also shed light on the external validity of the results from the main experimental sample. For example, it is possible that doctors have stronger relative concerns than the rest of the population due to their competitive profession. Also, because most doctors' incomes substantially exceed the subsistence level, they may care about positional externalities more than poor individuals. The observable characteristics of the online sample used in the auxiliary experiment also differ from those of the sample of medical students. For example, on average, participants in the auxiliary sample are older and less educated.[41] Moreover, the auxiliary sample may also differ in unobservable characteristics, such as their proclivity for status concerns.

7.2 Results

We recruited 1,245 U.S. respondents using Amazon Mechanical Turk. Table 5 replicates the preference estimation from Table 4, but using the auxiliary experiment instead of the main experiment. The comparison between Table 5 and Table 4 suggests that the results from the main residency match experiment are robust in this auxiliary experiment.

The first row of Table 5 corresponds to the baseline estimates, which uses the experimental and non-experimental variations in beliefs. The coefficients β^{ER} and β^{COL} are similar between the main experiment and the auxiliary experiment. Focusing on the entire sample, the estimated β^{ER} and β^{COL} are 1.141 and -1.262 in the main experiment (p-value=0.048 and p-value=0.017), and 1.293 and -1.962 in the auxiliary experiment (p-value<0.001 for both). That is, the coefficients have the same signs and similar magnitudes. The coefficients are

[39] Details about the recruitment are presented in Appendix D.14.

[40] The results from the previous section suggests that using hypothetical choices may not be as problematic as generally thought, to the extent that preferences inferred from expected choices are consistent with preferences inferred from actual choices.

[41] Descriptive statistics comparing sample characteristics between the main and auxiliary experiment are presented in Appendix Table D.10.

more precisely estimated in the auxiliary sample, in part due to the larger sample size.

To compare magnitude of relative income effects across the two experiments, we can compare the ratio $-\frac{\beta^{ER}}{\beta^{COL}}$, which corresponds to the marginal rate of substitution between relative income and cost of living. In the main experiment, the ratio $-\frac{\beta^{ER}}{\beta^{COL}}$ is 0.90 (s.e. 0.64). In the auxiliary experiment, we find a corresponding ratio of 0.66 (s.e. 0.20). That is, the auxiliary experiment suggests slightly weaker preferences for relative income than the main experiment, but that difference is statistically insignificant. The estimated preferences for relative income are similar across the two samples despite large observable differences in observable characteristics. This constitutes suggestive evidence that medical students are not special in terms of their preferences for relative income.

The second row presents the experimental estimates based on the variation driven by the source-randomization. The results from the auxiliary experiment are even more robust than the results from the main experiment. In the main experiment, the baseline estimates for β^{ER} are similar to the experimental estimates, and this is true again in the auxiliary experiment. In the main experiment, the experimental estimates for β^{COL} are statistically insignificant and smaller in magnitude than the baseline estimates. In the auxiliary sample, the experimental estimates for β^{COL} are negative, precisely estimated, and statistically significant at the 1% level.

Another finding from the main experiment is the heterogeneity in β^{ER} by relationship status. In the auxiliary experiment, we find evidence consistent with this heterogeneity, although it is less extreme. Table 5 shows that, when we break down β^{ER} by relationship status, the coefficient of β^{ER} is smaller among singles than among non-singles. This difference is economically large for the experimental estimates: β^{ER} is 2.578 for singles and 0.664 for non-singles. However, there are two notable differences: the difference is statistically insignificant in the auxiliary experiment (p-value=0.244), and less pronounced in magnitude than the corresponding heterogeneity in the main experiment. This difference between the two experiments can be attributed to the differences in the characteristics of singles across the two samples. Compared to singles in the auxiliary sample (i.e., people recruited randomly from an online job marketplace), singles in the medical student sample are more likely to be in prime dating age and less likely to have children from previous relationships. Therefore, it is likely that single medical students are more interested in their dating prospects.

8 Interpretation

8.1 Addressing Confounding Factors

When interpreting the coefficient β^{ER}, one potential concern is that individuals do not prioritize their relative income *per se*, but they use it as a signal for other city attributes that they do prioritize. However, given that medical students are highly informed and are making a high-stakes decision, this possibility seems unlikely. These students devote their entire fourth year of medical school to the Match. After the hospital visits, they have about two months to finalize their rankings. During this time, they continue to gather information to aid their decision. Because they have been to these locations and can easily obtain additional information directly, it is unlikely that they would rely on earnings rank to learn about other features of the locations. Nevertheless, we present direct tests for this confounding factor.

One version of this confounding factor is that individuals use information about relative income to make inferences about the expected cost of living. If participants in the NRMP believe that their earnings rank reflects the degree of competition with their neighbors for some goods, such as housing, it would be natural for them to learn about cost of living from information about their relative income. We can test this directly by examining how the information provision experiment affected posterior beliefs about cost of living.

In Figures 6 we use similar learning regressions from Section 5.3. The key difference is that the regressions in Section 5.3 explore the effect of relative income feedback on beliefs about relative income, whereas the ones in this section measure the effects of relative income feedback on beliefs about cost of living (and vice versa). Figure 6.a shows the effect of earnings rank feedback on posterior beliefs about the cost of living from the baseline survey (i.e., the short-term effect). The slope is close to zero (-0.003), precisely estimated (s.e. 0.006), and statistically insignificant. This coefficient suggests that increasing the observed earnings rank by 1 percentage point reduces their posterior beliefs about cost of living by 0.003 percentage points. To put this magnitude in context, Figure 5.d suggests that the effect of earnings rank on posterior beliefs about earnings rank is 0.873 (s.e. 0.011). The difference between this 0.873 and the -0.003 effect is economically large and statistically highly significant. Figure 6.b shows the effect of earnings rank feedback on beliefs about the cost of living from the follow-up survey (i.e., the long-term effect). Again, the effect is close to zero (-0.011), precisely estimated (s.e. 0.011), and statistically insignificant.

This evidence refutes the concern that subjects react to information about relative income because they infer something about the cost of living. As additional evidence that subjects see cost of living and relative income as two distinct features of the city, Figure 6.c and 6.d show that the converse also is true: feedback about cost of living does not affect short-term or long-term beliefs about relative income.

A second version of the confounding factor is that individuals use information about relative income to learn about other city characteristics, such as school quality and crime rates. Although these inferences would not be unreasonable, this confounding factor would likely bias our estimates in the opposite direction: if more affluent ponds tend to have desirable amenities, then individuals should prefer to live in more affluent ponds, which is the opposite of what we find.

We designed the auxiliary experiment to provide a direct test of this hypothesis. Towards the end of the auxiliary survey, after individuals received feedback about the cost of living and the relative income, we included a set of additional questions eliciting beliefs about other attributes of the two cities under consideration. We picked eight attributes that individuals could arguably perceive to be correlated to the average income in a city: quality of schools, crime rates, quality of health services, quality of public spaces, quality of the environment, quality of entertainment, share of college graduates, and share of supporters of Donald Trump.

One important difference in context is that, whereas subjects in the main experiment made a high-stakes decision for which they obtained substantial information, subjects in the auxiliary experiment had no incentives to be informed about the attributes of these cities. This difference increases the likelihood of the confounding factor: that is, because these auxiliary subjects do not actively gather information, it is reasonable for them to use information about earnings rank to make inferences about other unobserved city attributes. Thus, this confounding factor would play a smaller role in the main experiment than in the auxiliary experiment.

To measure the importance of this potential confounding factor, the third row of Table 5 estimates the experimental model, adding these eight perceptions as additional control variables. If individuals care about relative income because they learn about the other characteristics, the coefficient on β^{ER} should be zero after controlling for these additional perceptions.[42] On the contrary, the comparison between the second and third rows of Table 5 rejects the hypothesis of this confounding factor: controlling for the additional characteristics, if anything, increases the value of β^{ER}. For instance, among non-singles, the experimental estimate for β^{ER} is 2.578 (p-value=0.011) without these additional controls and 3.048 (p-value=0.004) with these additional controls. The difference between these two coefficients is statistically insignificant. The increase in β^{ER} caused by adding the extra controls is consistent with the previous argument that, if anything, this confounding factor leads to an underestimation of preferences for relative income.

[42]Consistent with the less informed nature of this subject pool, we find that the feedback about cost of living and earnings rank did affect a few of these additional beliefs. Results presented in Appendix D.11.

8.2 Favorite Interpretation

In light of this additional evidence, our favorite interpretation for the estimated preference for relative income among non-singles is that they care about relative income *per se*. That is, they see their neighbors' incomes as a negative externality. The literature on relative concerns provides several potential explanations for these preferences. Individuals may anticipate that their consumption aspirations will increase with the consumption of their neighbors (e.g., Frank 1985). Individuals could get a boost in happiness from observing that they are doing better than their neighbors (e.g., Boskin and Sheshinski 1978). They may prefer higher relative incomes, which allow them to obtain non-market goods and services (Cole et al. 1998), such as being invited on a date or to join a business venture or a club. Indeed, these preferences would be consistent with evidence that rich individuals get preferential treatment in a variety of social and professional interactions (Doob and Gross 1968; Fennis 2008; Nelissen and Meijers 2011) and evidence that individuals overspend on highly visible goods to appear rich to their peers (Heffetz 2011; Bursztyn et al. 2017a).

If the preferences for relative income reflect positional externalities, as in Luttmer (2005), it is then useful to compare the magnitude of our coefficients to those from Luttmer (2005).[43] The key specification from Luttmer (2005), which is estimated on the sample of non-single individuals, implies that most of the utility from consumption derives from relative consumption rather than absolute consumption: non-single individuals are willing to give up 1% of absolute consumption to decrease the median consumption of neighbors by 0.22%.[44] According to our baseline estimates from column (1) of Table 4, non-single individuals are willing to give up 1% of their absolute consumption to decrease the median consumption of their peers by 0.91% (90% confidence interval: [−0.18%, 2.05%]).[45] Compared to Luttmer (2005), our baseline estimates suggest a weaker role for relative concerns; however, this difference is not statistically significant. Compared to Luttmer (2005), the results from our auxiliary experiment also suggest a weaker role of relative concerns, with a statistically significant difference: our auxiliary estimates suggest that subjects are willing to give up 1% of absolute consumption to decrease the median consumption of their peers by 2.79% (90% confidence interval: [1.13%, 4.52%]).[46] Assuming that the estimates from Luttmer (2005) reflect real external-

[43]We focus on Luttmer (2005) because it uses data for the United States and is then the most comparable sample. The results are similar when we compare our estimates to estimates from other papers using subjective data (reported in Appendix D.12).

[44]Appendix D.12 provides the details for this calculation.

[45]For the average individual in the sample, we would need to decrease the median earnings in the area by 0.88% to allow the individual to climb up 0.519 (= 1/1.928) percentage points in the earnings rank.

[46]Of course, part of the difference may be due to differences in the subject pools: i.e., senior medical students having weaker preferences for relative concerns than the average U.S. resident. Also, any differences in the trade-offs measures with happiness and choice data would not imply that one of the two results are

ities, they suggest that individuals anticipate, at least partially, the negative externalities from richer neighbors.[47]

Last, our favorite interpretation for the heterogeneity in preferences for relative income by relationship status is due to dating prospects: being single increases the attractiveness of more affluent ponds, because rich peers make more desirable partners (Fisman et al. 2006). This results is also consistent with Gautier et al. (2010), who show that singles are willing to pay higher housing prices to benefit from a denser dating market in cities. Consistent with our findings for non-singles, they find that after getting married, the dating market benefits no longer matter for them, and couples move out of the city.

9 Conclusions

We present results from a survey of 1,100 medical students who participated in the National Resident Matching Program. These data provide unique revealed-preference evidence that, when choosing where to live, individuals care about their relative income. Furthermore, we find that individuals can differ substantially in their preferences for relative income: non-single individuals want to live in less affluent ponds, whereas single individuals prefer to live in more affluent ponds. Moreover, we present evidence that these preferences arise because individuals care about relative income *per se*.

A first avenue for future research is to find other contexts in which this revealed-preference method could be used to estimate preferences for relative income. For instance, although the settings may not be as clear as they are for the medical residency, multiple job markets require job seekers to choose between offers in different cities. Using a broader subject pool will help generalize the findings from this study and provide more data to study heterogeneity in preferences.

Future research also should investigate the mechanisms underlying individuals' concerns about relative income. For instance, there is little evidence as to whether relative concerns respond to instrumental motives (e.g., dating prospects) and non-instrumental motives (e.g., envy).[48] Similarly, there is little evidence on whether individuals care about their own per-

wrong: e.g., it is possible that the happiness estimates reflect the true extent to which people care about relative concerns, but when deciding where to live, individuals under-estimate how much their well-being will depend on relative consumption. Furthermore, we would need the standard errors from Luttmer (2005) to directly compare with our estimates.

[47]For a more direct comparison between happiness and choice data, we can also exploit the survey responses on expected happiness. We find that the marginal rates of substitution inferred by happiness are statistically indistinguishable from the marginal rates of substitution inferred by choice; however, due the lack of precision of the happiness estimates, we cannot rule out large discrepancies. Results presented in Appendix Table D.9.

[48]For instance, Cullen and Pakzad-Hurson (2017) show suggestive evidence that, in the context of an online work platform, concerns for relative wages operate though the instrumental channel.

ceptions of relative income (i.e., self image) or about how they are perceived by others (i.e., social image).[49] These additional hypotheses can be explored by using the same empirical framework proposed in this paper, but with additional treatment arms designed to test specific mechanisms.

References

Agarwal, N. (2015). An empirical model of the medical match. *The American Economic Review*, 105(7):1939–1978.

Albouy, D. (2008). Are big cities bad places to live? estimating quality of life across metropolitan areas. *NBER Working Paper No. 14472*.

Albouy, D. (2016). What are cities worth? land rents, local productivity, and the total value of amenities. *Review of Economics and Statistics*, 98(3):477–487.

Andrew E. Clark, C. S. and Yamada, K. (2017). When experienced and decision utility concur: The case of income comparisons. *Journal of Behavioral and Experimental Economics*, 70(Supplement C):1 – 9.

Armantier, O., Nelson, S., Topa, G., van der Klaauw, W., and Zafar, B. (2016). The price is right: Updating inflation expectations in a randomized price information experiment. *Review of Economics and Statistics*, 98(3):503–523.

Benjamin, D. J., Heffetz, O., Kimball, M. S., and Rees-Jones, A. (2012). What do you think would make you happier? what do you think you would choose? *The American Economic Review*, 102(5):2083–2110.

Benjamin, D. J., Heffetz, O., Kimball, M. S., and Rees-Jones, A. (2014). Can marginal rates of substitution be inferred from happiness data? evidence from residency choices. *The American Economic Review*, 104(11):3498–3528.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29(4):1165–1188.

Bertrand, M., Kamenica, E., and Pan, J. (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics*, 130(2):571–614.

Boskin, M. J. and Sheshinski, E. (1978). Optimal redistributive taxation when individual welfare depends upon relative income. *The Quarterly Journal of Economics*, 92(4):589–601.

[49] One exception is Bursztyn et al. (2017a) which shows suggestive evidence that, in the context of demand for premium credit cards, at least some of the conspicuous consumption operates through the self-image channel.

Bursztyn, L., Ferman, B., Fiorin, S., Kanz, M., and Rao, G. (2017a). Status goods: experimental evidence from platinum credit cards.

Bursztyn, L., Fujiwara, T., and Pallais, A. (2017b). "acting wife": Marriage market incentives and labor market investments. *The American Economic Review*, 107(11):3288–3319.

Cavallo, A., Cruces, G., and Perez-Truglia, R. (2017). Inflation expectations, learning, and supermarket prices: Evidence from survey experiments. *American Economic Journal: Macroeconomics*, 9(3):1–35.

Clark, A. E. and Oswald, A. J. (1996). Satisfaction and comparison income. *Journal of Public Economics*, 61(3):359–381.

Clark, A. E., Westergård-Nielsen, N., and Kristensen, N. (2009). Economic satisfaction and income rank in small neighbourhoods. *Journal of the European Economic Association*, 7(2-3):519–527.

Cole, H. L., Mailath, G. J., and Postlewaite, A. (1998). Class systems and the enforcement of social norms. *Journal of Public Economics*, 70(1):5–35.

Couture, V. and Handbury, J. (2016). Urban revival in america, 2000 to 2010. *University of Pennsylvania. Mimeo.*

Cruces, G., Perez-Truglia, R., and Tetaz, M. (2013). Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. *Journal of Public Economics*, 98:100–112.

Cullen, Z. B. and Pakzad-Hurson, B. (2017). Equilibrium effects of pay transparency. *HBS Working Paper No. 52648.*

Daniel, C. and O'Brien, M. (2008). Why study medicine? *The Student Doctor Network*, April 24.

Doob, A. N. and Gross, A. E. (1968). Status of frustrator as an inhibitor of horn-honking responses. *The Journal of Social Psychology*, 76(2):213–218.

Easterlin, R. A. (1974). Does economic growth improve the human lot? some empirical evidence. *Nations and households in economic growth*, 89:89–125.

Fennis, B. M. (2008). Branded into submission: Brand attributes and hierarchization behavior in same-sex and mixed-sex dyads. *Journal of Applied Social Psychology*, 38(8):1993–2009.

Ferrer-i Carbonell, A. (2005). Income and well-being: an empirical analysis of the comparison income effect. *Journal of Public Economics*, 89(5):997–1019.

Fisman, R., Iyengar, S. S., Kamenica, E., and Simonson, I. (2006). Gender differences in mate selection: Evidence from a speed dating experiment. *The Quarterly Journal of Economics*, 121(2):673–697.

Frank, R. H. (1985). The demand for unobservable and other nonpositional goods. *The American Economic Review*, 75(1):101–116.

Gautier, P. A., Svarer, M., and Teulings, C. N. (2010). Marriage and the city: Search frictions and sorting of singles. *Journal of Urban Economics*, 67(2):206 – 218.

Heffetz, O. (2011). A test of conspicuous consumption: Visibility and income elasticities. *Review of Economics and Statistics*, 93(4):1101–1117.

Hitsch, G., Hortacsu, A., and Ariely, D. (2010). What makes you click: An empirical analysis of online dating. *Quantitative Marketing and Economics*, 8(4):393–427.

Johansson-Stenman, O., Carlsson, F., and Daruvala, D. (2002). Measuring future grandparents' preferences for equality and relative standing. *The Economic Journal*, 112(479):362–383.

Kahneman, D. and Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive psychology*, 3(3):430–454.

Karadja, M., Mollerstrom, J., and Seim, D. (2017). Richer (and holier) than thou? the effect of relative income improvements on demand for redistribution. *Review of Economics and Statistics*, 99(2).

Kuziemko, I., Buell, R. W., Reich, T., and Norton, M. I. (2014). Last-place aversion: Evidence and redistributive implications. *The Quarterly Journal of Economics*, 129(1):105–149.

Luttmer, E. F. (2005). Neighbors as negatives: relative earnings and well-being. *The Quarterly Journal of Economics*, 120(3):963–1002.

Nelissen, R. M. and Meijers, M. H. (2011). Social benefits of luxury brands as costly signals of wealth and status. *Evolution and Human Behavior*, 32(5):343–355.

NRMP (2015). Data release and research committee: Results of the 2015 nrmp applicant survey by preferred specialty and applicant type. *National Resident Matching Program, Washington, DC. 2015.*

Perez-Truglia, R. (2016). The effects of income transparency on well-being: evidence from a natural experiment. *Mimeo.*

Perez-Truglia, R. (2017). Political conformity: Event-study evidence from the united states. *Review of Economics and Statistics*, forthcoming.

Perez-Truglia, R. and Cruces, G. (2017). Partisan interactions: Evidence from a field experiment in the united states. *Journal of Political Economy*, 125(4):1208–1243.

Rees-Jones, A. (2017). Suboptimal behavior in strategy-proof mechanisms: Evidence from the residency match. *Games and Economic Behavior*, In Press.

Rees-Jones, A. and Skowronek, S. (2017). Why do we lie in incentive-compatible mechanisms? evidence from the residency match. *Mimeo*.

Richins, M. L. and Dawson, S. (1992). A consumer values orientation for materialism and its measurement: Scale development and validation. *Journal of Consumer Research*, 19(3):303–316.

Roth, A. E. and Peranson, E. (1999). The redesign of the matching market for american physicians: Some engineering aspects of economic design. *The American Economic Review*, 89(4):748.

Roth, C. (2015). Conspicuous consumption and peer effects: Evidence from a randomized field experiment. *Mimeo*.

Senik, C. (2004). When information dominates comparison: Learning from russian subjective panel data. *Journal of Public Economics*, 88(9):2099–2123.

Shigeoka, H. and Yamada, K. (2016). Income-comparison attitudes in the us and the uk: Evidence from discrete-choice experiments. *NBER Working Paper No. 21998*.

Smither, R. D. and Houston, J. M. (1992). The nature of competitiveness: The development and validation of the competitiveness index. *Educational and Psychological Measurement*, 52(2):407–418.

Solnick, S. J. and Hemenway, D. (1998). Is more always better?: A survey on positional concerns. *Journal of Economic Behavior & Organization*, 37(3):373–383.

Tideman, S., Frijters, P., and Shields, M. A. (2008). Relative income, happiness, and utility: An explanation for the easterlin paradox and other puzzles. *Journal of Economic literature*, 46(1):95–144.

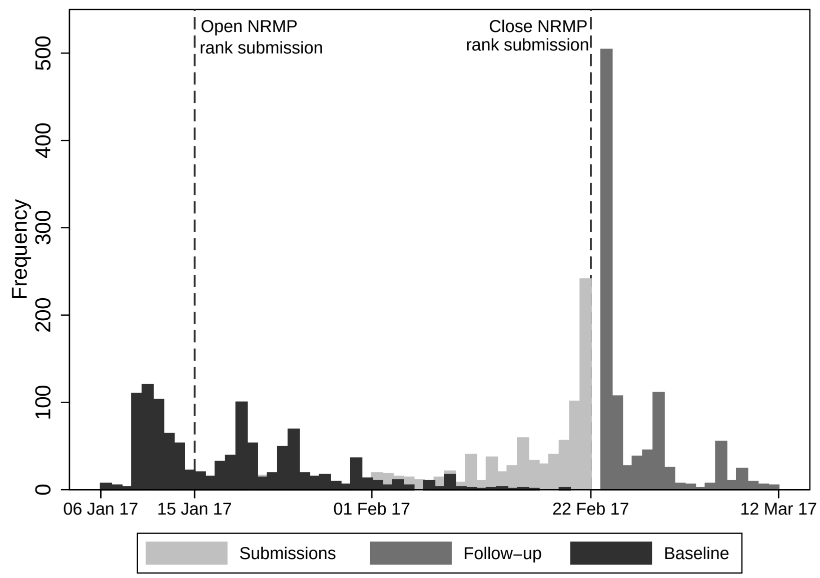
Van de Stadt, H., Kapteyn, A., and Van de Geer, S. (1985). The relativity of utility: Evidence from panel data. *The Review of Economics and Statistics*, 67(2):179–187.

Van Gelder, M. M., Bretveld, R. W., and Roeleveld, N. (2010). Web-based questionnaires: the future in epidemiology? *American Journal of Epidemiology*, 172(11):1292–1298.

Wiswall, M. and Zafar, B. (2014). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, 82(2):791–824.

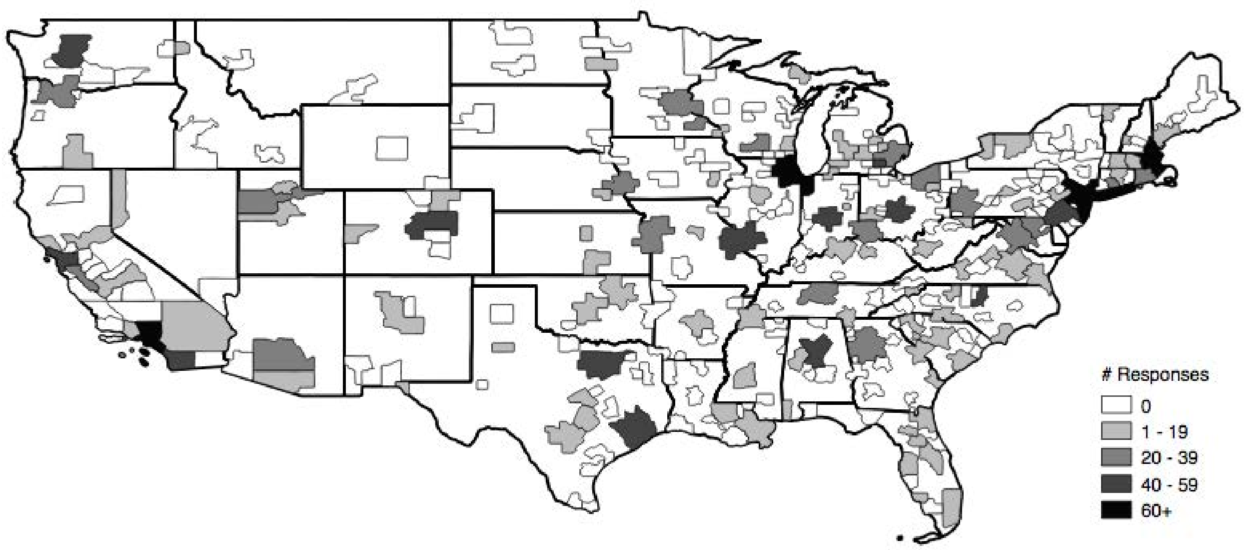
Yamada, K. and Sato, M. (2016). *Another avenue for anatomy of income comparisons: Evidence from hypothetical choice experiments.* Behavioral Economics of Preferences, Choices, and Happiness; Springer.

Figure 1: Distribution Over Time of Survey Responses and NRMP Rank Submissions



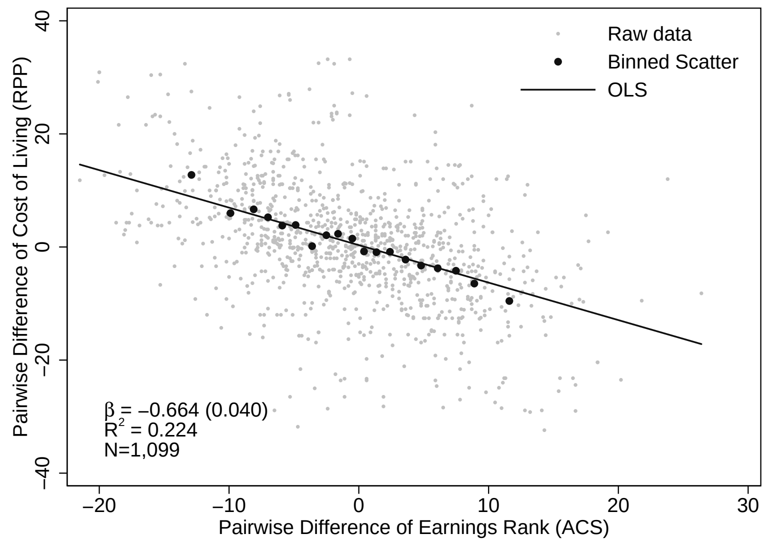
Notes: Distribution of timing of responses to Baseline and Follow-up Surveys, and NRMP rank submission dates (as reported by respondents in the follow-up survey).

Figure 2: Geographic Distribution of Choice Set



Notes: Geographical distribution of metropolitan areas where top-2 residency programs of respondents are located, for the continental United States. No responses were located in Hawaii, while Alaska only has 2 responses. Only metropolitan areas with a residency program participating in the 2017 NRMP are displayed (279 in total).

Figure 3: Variation in Cost of Living and Earnings Rank

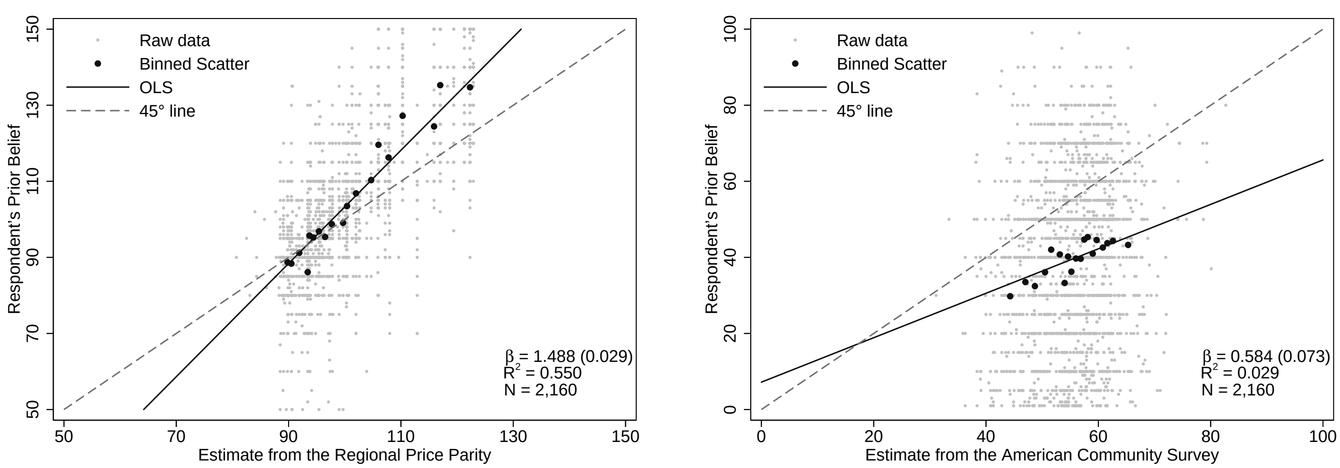


Notes: The gray dots correspond to the raw scatterplot, and the darker dots correspond to the binned-scatterplot based on 20 bins. Slopes (β, with robust standard errors in parentheses) and R^2 are based on a linear regression. All variables for x-axis and y-axis correspond to pairwise differences across the two cities that the subject is considering submitting to the algorithm. Data from survey responses, the Regional Price Parity Index (for cost of living) and the American Community Survey (for earnings rank).

Figure 4: Comparison Between Prior Beliefs and Statistics

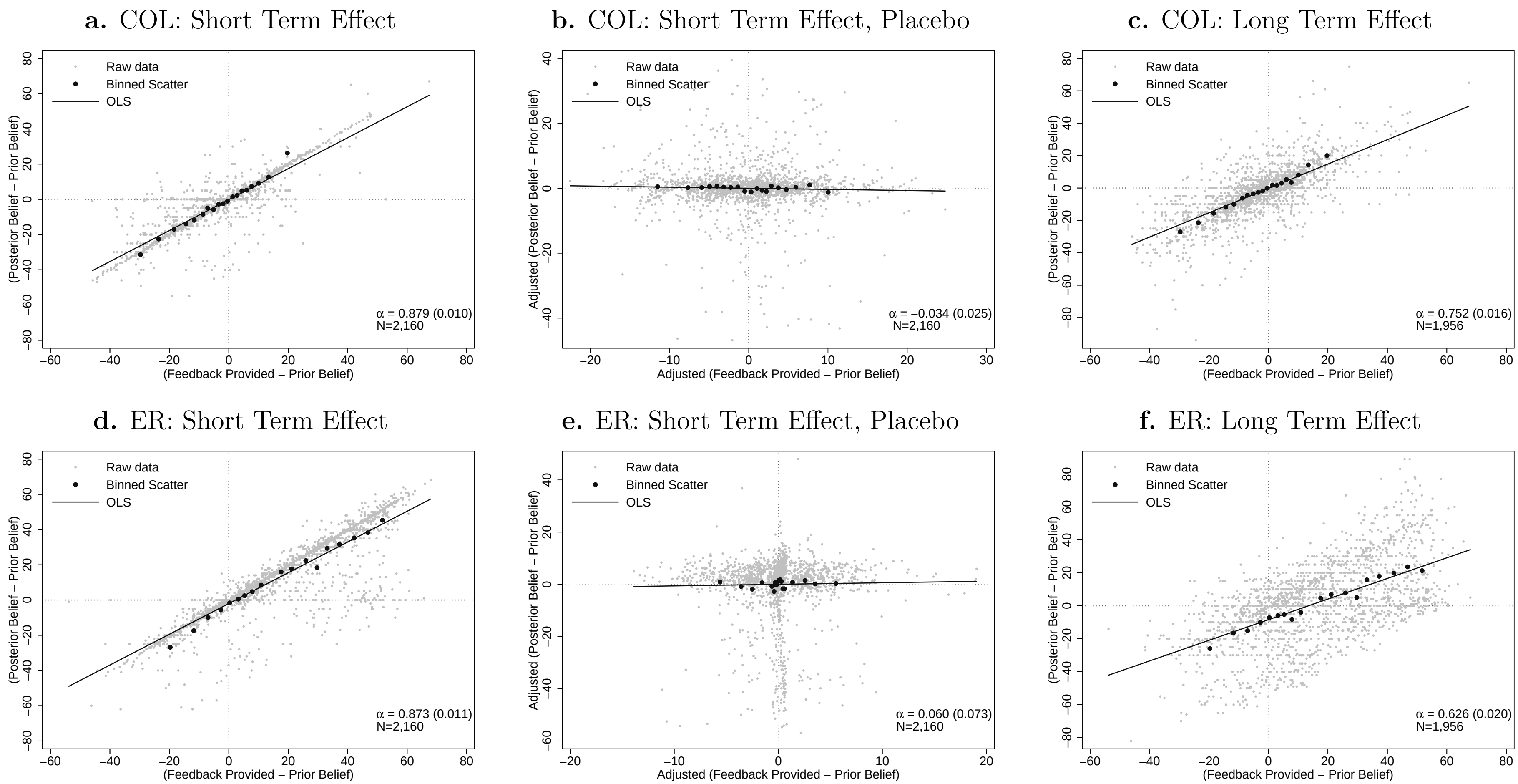
a. Cost of Living, in Levels

b. Earnings Rank, in Levels



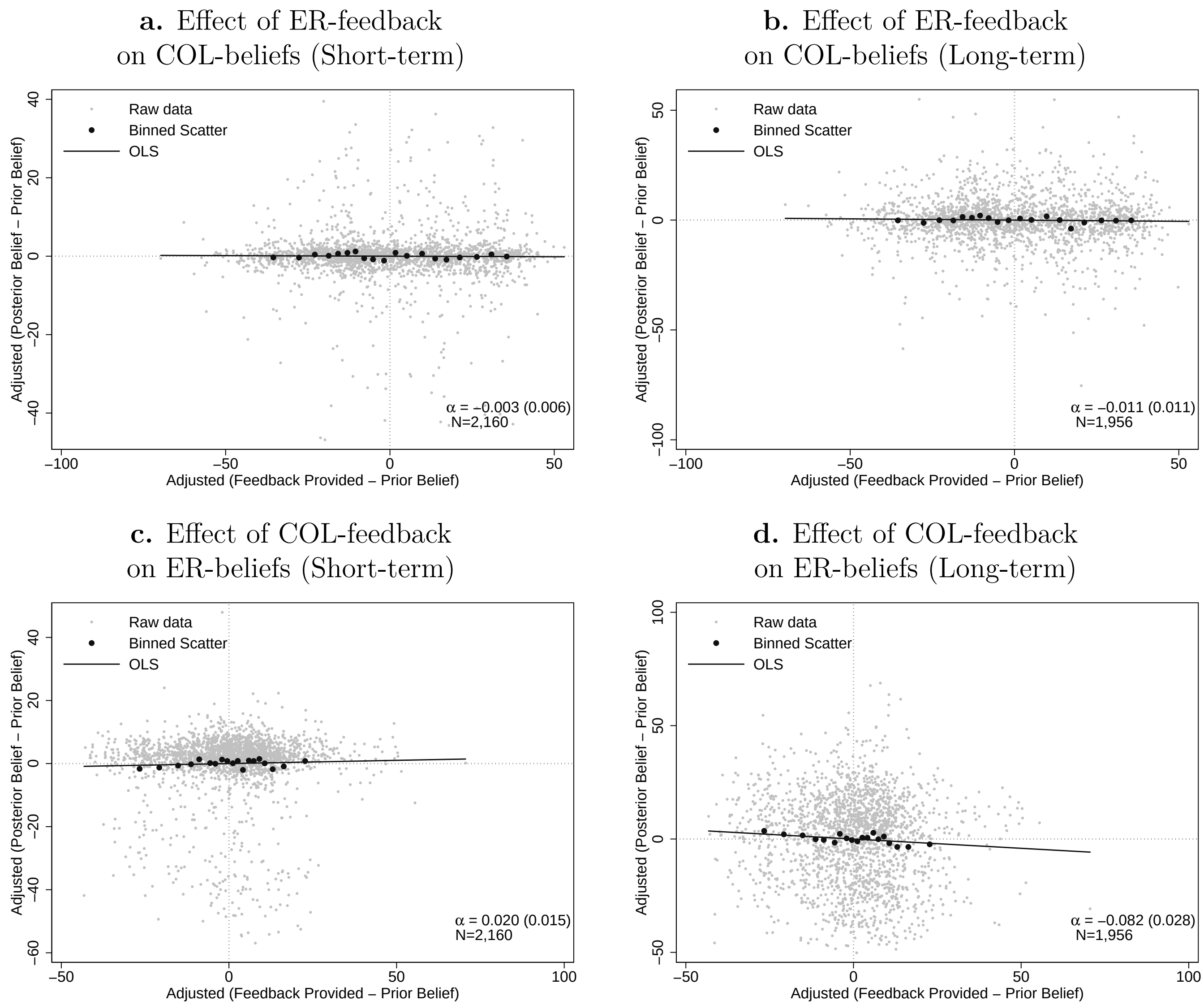
Notes: Comparison between respondent's perceptions before the information provision (i.e., prior beliefs) and statistics. The gray dots correspond to the raw scatterplot, and the darker dots correspond to the binned-scatterplot based on 20 bins. Panels a and b present data in levels (i.e., two observations per individual, one for each of their options). The slope (β, with robust standard errors in parentheses) and R^2 are based on a linear regression.

Figure 5: Learning from the Experimental Feedback



Notes: Comparison between the difference in statistics and respondent's perceptions before the information provision (i.e., prior beliefs), and difference in respondent's perceptions after the information provision (i.e., posterior beliefs) and prior beliefs. The gray dots correspond to the raw scatterplot, and the darker dots correspond to the binned-scatterplot based on 20 bins. Panel b and e shows a placebo test where we compare the difference between the alternative feedback and prior belief to the difference between the posterior and prior beliefs, adjusting for the shown statistic. Panel c and f uses respondent's perceptions measured in the follow-up survey as posterior belief. The slope (α, with robust standard errors in parentheses) is based on a linear regression.

Figure 6: Effect of Earning Rank Feedback on Posterior Belief on Cost of Living (and vice-versa)



Notes: Comparison between the difference in statistics and respondent's perceptions before the information provision (i.e., prior beliefs), and difference in respondent's perceptions after the information provision (i.e., posterior beliefs) and prior beliefs. The gray dots correspond to the raw scatterplot, and the darker dots correspond to the binned-scatterplot based on 20 bins. Panels shows the extent to which respondents adjust their perceptions on earnings rank (cost of living) as a result in their perception gap in cost of living (earnings rank) adjusted for the perceptions gap in ER (COL). The slope (α, with robust standard errors in parentheses) is based on a linear regression.

Table 1: Descriptive Statistics and Randomization Balance

	All	RPP; ACS	RPP; CPS	COLI; ACS	COLI; CPS	F-test P-value
Male (=1)	0.481	0.452	0.491	0.481	0.502	0.688
	(0.015)	(0.030)	(0.031)	(0.031)	(0.030)	
Age	27.091	27.092	27.104	26.985	27.181	0.863
	(0.083)	(0.164)	(0.165)	(0.145)	(0.187)	
Nr Kids	0.132	0.125	0.164	0.104	0.137	0.553
	(0.014)	(0.027)	(0.033)	(0.026)	(0.029)	
Single (=1)	0.354	0.401	0.312	0.343	0.358	0.189
	(0.015)	(0.030)	(0.028)	(0.029)	(0.029)	
Dual Match (=1)	0.074	0.077	0.059	0.104	0.055	0.157
	(0.008)	(0.016)	(0.014)	(0.019)	(0.014)	
US News Rank	58.81	58.849	59.104	58.604	58.683	0.996
	(0.787)	(1.612)	(1.560)	(1.568)	(1.565)	
Prior: $COL^i_{1,2}$	0.409	0.445	-0.238	-0.567	1.982	0.506
	(0.640)	(1.364)	(1.134)	(1.308)	(1.303)	
Prior: $ER^i_{1,2}$	0.394	0.162	0.71	-0.526	1.221	0.595
	(0.467)	(0.903)	(0.925)	(0.906)	(0.999)	
Observations	1,080	272	269	268	271	

Notes: Individual characteristics obtained from baseline survey. Column (1) corresponds to all respondents, and columns (2) through (4) correspond to each of the four treatment groups given by all the possible combinations from the source-randomization experiment. RPP and COLI are the two sources used to compute the cost of living feedback (corresponding to the Regional Price Parity Index and the Cost of Living Index, respectively). ACS and CPS are the two sources used to compute the earnings ranking feedback (corresponding to the American Community Survey and the Current Population Survey, respectively). The final column presents p-value for test of the null hypothesis that the mean characteristic is equal across all four treatment groups. All variables constructed from the survey data, except for the U.S. News Rank which was taken from the U.S. News rank of medical schools for 2016.

Table 2: Preference for Relative Income: Baseline Estimates

		By Relationship Status		By Gender		By Specialty Salary	
	All	Non-Single	Single	Female	Male	$\leq$ \$229,000	$>$ \$229,000
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
β^{ER}	0.995*	2.236***	-1.538*	1.041	0.896	1.433*	0.777
	(0.539)	(0.669)	(0.880)	(0.755)	(0.781)	(0.732)	(0.797)
β^{COL}	-1.073**	-1.087	-1.058	-0.972	-1.443*	-0.690	-1.238*
	(0.485)	(0.663)	(0.749)	(0.679)	(0.753)	(0.713)	(0.690)
Diff. P-value *[q-value]:*							
ER		0.001 *[0.030]*		0.894 *[0.974]*		0.544 *[0.954]*	
COL		0.977 *[0.977]*		0.642 *[0.954]*		0.580 *[0.954]*	
Pseudo R^2	0.025	0.047	0.026	0.043	0.032	0.033	0.032
Observations	1,080	698	382	560	520	549	531

Notes: Heteroskedasticity-robust standard errors in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Raw Probit coefficients. Each column corresponds to a different Probit regression of expected rank order submission on posterior beliefs about cost of living and earnings rank, from the baseline survey, including the baseline controls listed in section 3. Columns (2) through (7) show estimates when restricting sample to different subgroups: columns 2 and 3 by Non-Single (i.e., married or in a long-term relationship) or Single, columns 4 and 5 by gender, columns 5 and 7 by expected salary of specialty after residency (over and below the median value of \$229,000). P-values corresponds to the test of the null hypothesis that the coefficients are equal between the two sub-groups, multiple-testing q-values based on Benjamini and Yekutieli (2001) presented in brackets.

Table 3: Preference for Relative Income: Robustness to Alternative Control Variables

	Panel A: β^{ER}			Panel B: β^{COL}			Pseudo R^2		
	Non-Single (1)	Single (2)	All (3)	Non-Single (4)	Single (5)	All (6)	Non-Single (7)	Single (8)	All (9)
No Controls	1.961***	-1.480*	0.873*	-0.812	-1.131*	-0.894**	0.032	0.017	0.015
	(0.663)	(0.841)	(0.531)	(0.523)	(0.589)	(0.382)			
Baseline	2.236***	-1.538*	0.995*	-1.087	-1.058	-1.073**	0.047	0.026	0.025
	(0.669)	(0.880)	(0.539)	(0.663)	(0.749)	(0.485)			
Demographic	2.288***	-0.871	1.176**	-1.219*	-1.712**	-1.342***	0.055	0.045	0.031
	(0.715)	(0.977)	(0.578)	(0.628)	(0.713)	(0.468)			
Amenities	2.056***	-1.381	0.958*	-0.718	-1.265	-0.898*	0.037	0.022	0.018
	(0.669)	(0.853)	(0.538)	(0.630)	(0.816)	(0.481)			
Geography	2.064***	-1.551	1.001*	-1.626**	-1.783***	-1.572***	0.059	0.054	0.039
	(0.733)	(1.004)	(0.593)	(0.652)	(0.646)	(0.461)			
Economic	1.914***	-1.191	0.946*	-0.467	-1.647**	-0.868*	0.036	0.047	0.019
	(0.684)	(0.941)	(0.566)	(0.670)	(0.812)	(0.498)			
State Dummies	2.901***	-1.943**	1.084*	-1.090	-1.219	-0.968*	0.105	0.149	0.049
	(0.703)	(0.907)	(0.555)	(0.671)	(0.939)	(0.502)			
Obj. Program Chars.	1.987***	-1.451*	0.919*	-0.868*	-1.141*	-0.912**	0.037	0.019	0.018
	(0.686)	(0.842)	(0.540)	(0.527)	(0.597)	(0.386)			
Subj. Program Chars.	2.222***	-1.320	1.199**	-1.210**	-1.678***	-1.277***	0.142	0.137	0.123
	(0.730)	(1.040)	(0.605)	(0.587)	(0.619)	(0.425)			
All Controls	2.221**	-3.177*	0.703	-0.295	-5.067***	-0.862	0.285	0.492	0.218
	(0.941)	(1.854)	(0.700)	(1.083)	(1.712)	(0.704)			

Notes: Heteroskedasticity-robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Raw Probit coefficients. Each row corresponds to a separate regression of expected rank on posterior beliefs about cost of living and earnings rank, from the baseline survey. All regressions include as controls the log difference in nominal income and a constant. The first row does not include any additional controls. The second row includes the baseline controls listed in section 3. The third to last rows use different sets of additional controls, listed in section 6.3. Results are based on 1,080 individual responses (698 from non-singles and 382 from singles), except for the last row, which is restricted to the follow-up sample (978 responses, 595 from non-singles and 311 from singles).

Table 4: Preference for Relative Income: Experimental Estimates

	Panel A: β^{ER}			Panel B: β^{COL}		
	Non-Single (1)	Single (2)	All (3)	Non-Single (4)	Single (5)	All (6)
Baseline	2.380***	-1.656*	1.141**	-1.234*	-1.379*	-1.262**
	(0.702)	(0.991)	(0.577)	(0.743)	(0.772)	(0.531)
Experimental	2.977**	-4.964**	0.867	0.353	1.663	0.662
	(1.331)	(1.974)	(1.151)	(1.160)	(1.286)	(0.881)
Experimental, Long Term	1.993*	-5.285***	-0.029	1.662*	0.251	1.012
	(1.188)	(1.984)	(1.071)	(1.005)	(1.359)	(0.821)
Experimental, Falsification	-0.007	0.040	0.004	0.037	0.021	0.031
	(0.998)	(1.732)	(0.837)	(0.855)	(1.123)	(0.651)

Notes: Heteroskedasticity-robust standard errors in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Raw Probit (or IV-Probit) coefficients restricting sample to respondents who completed the follow-up survey. All regressions include the baseline controls listed in section 3. The independent variables are the posterior beliefs about cost of living and earnings rank, from the baseline specification. The first row corresponds to the baseline Probit specification. The second through third row correspond to IV-Probit regressions, using the variation in perceptions generated by the source-randomization experiment as instrumental variables. The first and second rows use the expected rank order submission (from the baseline survey) as dependent variable. The third row uses the final rank order submission (from the follow-up survey) as dependent variable. The fourth row provides corresponds to a falsification test that uses the same IV-Probit specification from the second row, but using the list order (i.e., the order in which programs are listed at the beginning of the survey) as dependent variable instead of the rank order. To estimate this IV-Probit model, we randomly assign programs to be program 1 and program 2, and then we use as dependent variable a dummy that takes the value 1 if program 1 was listed first at the beginning of the survey. We repeat this procedure 1,000 times and report the average and standard error from the distribution of coefficients. All results based on the sample of individuals who completed the follow-up survey (978 responses, 647 from non-singles and 311 from singles).

Table 5: Results from Auxiliary Experiment: Preference for Relative Income

	Panel A: β^{ER}			Panel B: β^{COL}		
	Non-Single (1)	Single (2)	All (3)	Non-Single (4)	Single (5)	All (6)
Baseline	1.408***	1.095**	1.293***	-2.203***	-1.618***	-1.962***
	(0.376)	(0.478)	(0.292)	(0.463)	(0.566)	(0.364)
Experimental	2.578**	0.664	1.706**	-2.385***	-2.956***	-2.528***
	(1.019)	(1.272)	(0.816)	(0.666)	(0.917)	(0.531)
Experimental, Additional Controls	3.048***	0.452	1.902**	-2.329***	-3.753***	-2.688***
	(1.064)	(1.430)	(0.872)	(0.691)	(0.906)	(0.563)

Notes: Heteroskedasticity-robust standard errors in parenthesis. All regressions include the baseline controls listed in section 3 with the exception of program characteristics. The independent variables are the posterior beliefs about cost of living and earnings rank, from the baseline specification. The first row corresponds to the baseline Probit specification. The second through third row correspond to IV-Probit regressions, using the variation in perceptions generated by the source-randomization experiment as instrumental variables. The third row includes additional controls for differences in city perceptions: quality of schools, crime rates, quality of health, quality of public spaces, quality of the environment, quality of entertainment, quality of colleges, fraction voting Trump in presidential election. All results based on the sample of respondents in the United States on Amazon Mechanical Turk (1,245 responses, 829 from non-singles and 416 from singles).

Online Appendix: For Online Publication Only

A Snapshots of Invitations and Surveys

Here we include snapshots with a sample of the baseline survey (A.1) and follow-up survey (A.2). Additionally, this Appendix also includes a snapshot of an invitation sent out to the deans (A.3), the invitation sent out to the medical students to participate in the baseline survey (A.4), the invitation sent out to students inviting them to the follow-up survey (A.5), and a snapshot of the project's website.

A.1 Sample Questionnaire: Baseline Survey

This survey has the objective of understanding how participants of the 2017 NRMP make their ranking decisions. Even though it may not benefit you directly, the results from this survey may benefit the medical students participating in future years.

We anticipate that this survey will take between 8 to 10 minutes to complete. Eligible participants completing the entire survey will be paid $10 in the form of an Amazon Gift Card (note: you must have a .edu email address).

Your participation is voluntary, and is greatly appreciated. You may withdraw from the study at any time. Your responses will be used solely for research purposes and will be kept strictly confidential, used only by the Principal Investigators. For more details about this survey, including contact information, please visit the project's website.

To be eligible to participate in this survey, **you must be a medical student participating in the 2017 Main Residency Match and not yet submitted your rankings.**

- YES, I am participating in the 2017 Main Residency Match and would like to complete the survey

>>

NOTE: Please answer questions carefully, it is not possible to go back and change an answer.

Where are you attending Medical School?

State

Medical School

Which **match** will you be participating in?
(Note: this is referring to the match, not necessarily your specialty)

Main Residency (Opens Jan 15)

Will you register with the NRMP for a dual match?

- Yes
- No

Did you already submit your ranking to the NRMP?

- Yes
- No

In the next couple of weeks you will be submitting your rankings to the Main Residency Match.

Please tell us (in no particular order) the top two Residency Programs you are thinking about ranking in the Main Residency Match.

Enter information for first program.

State	California
Metropolitan Area	Los Angeles-Long Beach-Anaheim, CA
Program	UCLA Medical Center

Specialty:

Internal Medicine (IM)

What is the annual salary you are being offered here? (pre-tax, in dollars)

54000

>>

Enter information for second program.

State: Illinois

Metropolitan Area: Champaign-Urbana, IL

Program: Carle Foundation Hospital

Specialty:

Internal Medicine (IM)

What is the annual salary you are being offered here? (pre-tax, in dollars)

54000

>>

Now, we want to ask you a couple of questions about the two cities you are considering living in.

Let's start with the expected cost of living. You probably noticed that the average prices of goods and services are different across different cities. As a result, with the same income, you would be able to buy more things in some cities and less in other cities.

Imagine that you chose to work in the **Los Angeles-Long Beach-Anaheim, CA** metro area. Would you expect your cost of living in this city to be cheaper or more expensive than the U.S. average?

- ○ cheaper
- ● more expensive

How much more expensive is the Los Angeles-Long Beach-Anaheim, CA metro area than the U.S. average?

20%

>>

Imagine that you chose to work in the **Champaign-Urbana, IL** metro area. Would you expect your cost of living in this city to be cheaper or more expensive than the U.S. average?

- (●) cheaper
- ( ) more expensive

How much cheaper is the Champaign-Urbana, IL metro area than the U.S. average?

10%

>>

Now we want to ask you about your expected earnings ranking. This ranking is defined as the share of the working individuals of a city who earn less than you. You probably noticed that the distribution of earnings is different across different cities. As a result, with the same earnings, you may be relatively rich in some cities but relatively poor in other cities.

Imagine that you chose to work in **Los Angeles-Long Beach-Anaheim, CA**. With your individual annual earnings of **$ 54000**, you would be richer than what percentage of **Los Angeles-Long Beach-Anaheim, CA**'s individual earners?

Richer than 35% of individual earners

>>

7

Imagine that you chose to work in **Champaign-Urbana, IL**. With your individual annual earnings of **$ 54000**, you would be richer than what percentage of **Champaign-Urbana, IL**'s individual earners?

Richer than 47% of individual earners

>>

Now, we want to share some information with you, related to the characteristics of the two cities that you are considering living in. Please take a moment to review the information carefully.

Note: *this information is only shown once and you will not be able to come back to it.*

First, find below some estimates of the cost of living:

The **Los Angeles-Long Beach-Anaheim, CA** metro area is **17.0% more expensive** than the U.S. average.

The **Champaign-Urbana, IL** metro area is **6.6% cheaper** than the U.S. average.

Source: based on most recent data from the Bureau of Economic Analysis.

Second, find below some estimates of the earnings ranking:

With your individual annual earnings of **$ 54000**, you would be richer than **57.9%** of **Los Angeles-Long Beach-Anaheim, CA**'s population.

With your individual annual earnings of **$ 54000**, you would be richer than **60.3%** of **Champaign-Urbana, IL**'s population.

Source: based on most recent data from the American Community Survey.

That was all the information that we wanted to share with you. Now that you have reviewed this information, we would like to ask you again about your expected cost of living and earning rankings.

Let's start with the cost of living:

Imagine that you chose to work in the **Los Angeles-Long Beach-Anaheim, CA** metro area. Would you expect your cost of living in this city to be cheaper or more expensive than the U.S. average?

- ○ cheaper
- ● more expensive

How much more expensive is the Los Angeles-Long Beach-Anaheim, CA metro area than the U.S. average?

17%

>>

Imagine that you chose to work in the **Champaign-Urbana, IL** metro area. Would you expect your cost of living in this city to be cheaper or more expensive than the U.S. average?

- (●) cheaper
- ( ) more expensive

How much cheaper is the Champaign-Urbana, IL metro area than the U.S. average?

6%

>>

Imagine that you chose to work in **Los Angeles-Long Beach-Anaheim, CA**. With your individual annual earnings of **$ 54000**, you would be richer than what percentage of **Los Angeles-Long Beach-Anaheim, CA**'s individual earners?

Richer than 58% of individual earners

>>

Imagine that you chose to work in **Champaign-Urbana, IL**. With your individual annual earnings of **$ 54000**, you would be richer than what percentage of **Champaign-Urbana, IL**'s individual earners?

Richer than 60% of individual earners

>>

We understand this is a lot of information to process, so we will help you make the comparison simpler. According to your final answers about incomes, cost of living and relative earnings:

- If you chose to live in Los Angeles-Long Beach-Anaheim, CA, you would be able to afford 19.7% less than if you chose to live in Champaign-Urbana, IL.

- If you chose to live in Los Angeles-Long Beach-Anaheim, CA, your earnings ranking would be 3.3% lower than if you chose to live in Champaign-Urbana, IL.

As of this moment: of the two programs discussed so far, which one do you expect to **rank higher** for the NRMP?

- ○ Very likely UCLA Medical Center (Los Angeles-Long Beach-Anaheim, CA)
- ○ Likely UCLA Medical Center
- ○ Leaning UCLA Medical Center
- ○ Leaning Carle Foundation Hospital
- ○ Likely Carle Foundation Hospital
- ○ Very likely Carle Foundation Hospital (Champaign-Urbana, IL)

>>

16

If assigned to it, in which of the two programs would you expect to **live a happier life**?

- ○ Very likely UCLA Medical Center (Los Angeles-Long Beach-Anaheim, CA)
- ○ Likely UCLA Medical Center
- ○ Leaning UCLA Medical Center
- ○ Leaning Carle Foundation Hospital
- ○ Likely Carle Foundation Hospital
- ○ Very likely Carle Foundation Hospital (Champaign-Urbana, IL)

>>

To get a general picture of the people answering this survey, we would like to ask you a few things about yourself. Please remember that your answers are confidential and that your name is not collected as part of this study.

Please indicate your gender:

- ○ Female
- ○ Male

How old are you?

Age

What is your relationship status?

- ○ Single
- ○ In a long-term relationship
- ○ Married

How many children do you have?

Recent research on decision making shows that choices are affected by the context in

18

Recent research on decision making shows that choices are affected by the context in which they are made. Differences in how people feel, in their previous knowledge and experience, and in their environment can influence the choices they make. To help us understand how people make decisions, we are interested in information about you, specifically whether you actually take the time to read the instructions; if you don't, some results may fail to tell us very much about decision making in the real world. To help us confirm that you have read these instructions, please ignore the question below about how you are feeling and instead check only the "none of the above" option. Thank you very much.

☐ Interested	☐ Enthusiastic	☐ Inspired
☐ Distressed	☐ Proud	☐ Determined
☐ Excited	☐ Irritable	☐ Attentive
☐ Scared	☐ Alert	☐ None of the above

>>

Thank you so much for completing the survey! As a token of our appreciation, we want to send you a $10 Amazon Gift Card. Please note that you may only participate once.

We need your official university email address (.edu) to be able to: (i) email you the Amazon gift card; and (ii) verify that you are a medical student participating in the 2017 NRMP.

I certify that I am a medical student participating in the 2017 NRMP match.

Please sign with your university (.edu) email address:

As a reminder, your email address and survey responses will be kept strictly confidential.

>>

20

A.2 Sample Questionnaire: Follow-Up Survey

Thank you for volunteering to participate in our follow-up survey! Remember that your responses will be used solely for research purposes and will be kept strictly confidential. You may withdraw from the survey at any time.

We estimate that it will take you around 5 minutes to complete the survey. As a token of our appreciation, we will send you a $5 Amazon gift card for completing this survey.

For more details about the survey, including contact information, please visit the project's website.

>>

On what date did you submit your preference ranking to the 2017 Main Residency Match?

Feb 20

>>

In the initial survey you listed two of your favorite programs.

When you submitted your preference ranking to the 2017 Main Residency Match on Feb 20, which of these two programs did you **rank higher**?

- ○ UCLA Medical Center (Los Angeles-Long Beach-Anaheim, CA)
- ○ Carle Foundation Hospital (Champaign-Urbana, IL)

>>

3

If assigned to it, in which of the two programs would you expect to **live a happier life**?

- ○ Very likely UCLA Medical Center (Los Angeles-Long Beach-Anaheim, CA)
- ○ Likely UCLA Medical Center
- ○ Leaning UCLA Medical Center
- ○ Leaning Carle Foundation Hospital
- ○ Likely Carle Foundation Hospital
- ○ Very likely Carle Foundation Hospital (Champaign-Urbana, IL)

>>

Now, we want to ask you a couple of questions about the two cities where you may live. When you took the survey a month ago, we asked these same questions. We are asking them again to see if your perceptions have changed.

Let's start with the expected cost of living. You probably noticed that the average prices of goods and services are different across different cities. As a result, with the same income, you would be able to buy more things in some cities and less in other cities.

Imagine that you chose to work in the Los Angeles-Long Beach-Anaheim, CA metro area. Would you expect your cost of living in this city to be cheaper or more expensive than the U.S. average?

- ○ cheaper
- ● more expensive

How much more expensive is the Los Angeles-Long Beach-Anaheim, CA metro area than the U.S. average?

>>

5

Imagine that you chose to work in the Champaign-Urbana, IL metro area. Would you expect your cost of living in this city to be cheaper or more expensive than the U.S. average?

- ● cheaper
- ○ more expensive

How much cheaper is the Champaign-Urbana, IL metro area than the U.S. average?

>>

6

Now we want to ask you about your expected earnings ranking. This ranking is defined as the share of the working individuals of a city who earn less than you. You probably noticed that the distribution of earnings is different across different cities. As a result, with the same earnings, you may be relatively rich in some cities but relatively poor in other cities.

Imagine that you chose to work in Los Angeles-Long Beach-Anaheim, CA. With your individual annual earnings of **$** 54000, you would be richer than what percentage of Los Angeles-Long Beach-Anaheim, CA's individual earners?

Richer than 58% of individual earners

>>

Imagine that you chose to work in Champaign-Urbana, IL. With your individual annual earnings of **$** 54000, you would be richer than what percentage of Champaign-Urbana, IL's individual earners?

Richer than 60% of individual earners

>>

Now we want to ask you to compare other aspects of these two programs.

In which program do you expect to have a greater sense of purpose in life?

- Very likely UCLA Medical Center (Los Angeles-Long Beach-Anaheim, CA)
- Likely UCLA Medical Center
- Leaning UCLA Medical Center
- Leaning Carle Foundation Hospital
- Likely Carle Foundation Hospital
- Very likely Carle Foundation Hospital (Champaign-Urbana, IL)

Which program do you think will give you higher prestige and status?

- Very likely UCLA Medical Center (Los Angeles-Long Beach-Anaheim, CA)
- Likely UCLA Medical Center
- Leaning UCLA Medical Center
- Leaning Carle Foundation Hospital
- Likely Carle Foundation Hospital
- Very likely Carle Foundation Hospital (Champaign-Urbana, IL)

Which program do you think will give you better future career prospects?

- Very likely UCLA Medical Center (Los Angeles-Long Beach-Anaheim, CA)
- Likely UCLA Medical Center

Which program do you think will give you higher prestige and status?

- Very likely UCLA Medical Center (Los Angeles-Long Beach-Anaheim, CA)
- Likely UCLA Medical Center
- Leaning UCLA Medical Center
- Leaning Carle Foundation Hospital
- Likely Carle Foundation Hospital
- Very likely Carle Foundation Hospital (Champaign-Urbana, IL)

Which program do you think will give you better future career prospects?

- Very likely UCLA Medical Center (Los Angeles-Long Beach-Anaheim, CA)
- Likely UCLA Medical Center
- Leaning UCLA Medical Center
- Leaning Carle Foundation Hospital
- Likely Carle Foundation Hospital
- Very likely Carle Foundation Hospital (Champaign-Urbana, IL)

>>

10

From 1 (most important) to 5 (least important): How would you rank the following aspects of life? *(no ties)*

	1	2	3	4	5
Happiness	○	○	○	○	○
Health	○	○	○	○	○
Sense of purpose	○	○	○	○	○
Spirituality	○	○	○	○	○
Control over your life	○	○	○	○	○

>>

When you submitted your preference ranking to the 2017 Main Residency Match on Feb 20, how many programs did you rank in total?

>>

Now we want to ask you a few more questions about your background, your beliefs and your values.

Did you grow up in the United States?

- ○ Yes
- ○ No

>>

13

More precisely, in which of the following did you spend the most time while growing up?

State

Metro area

>>

14

Imagine that you face the following situation. You earn $50,000 per year and have an earnings ranking of 50% (that is, you earn more than 50% of the individuals living in your same city). Now consider the following two events:

EVENT A: The cost of living in this city decreases by 10%, so you and all other individuals in the city would be able to afford 10% more consumption. After this event, you think you would be:

- ○ Better off
- ○ Slightly better off
- ○ The same
- ○ Slightly worse off
- ○ Worse off

EVENT B: Your own income and your own cost of living do not change, so your own consumption stays the same. However, all other individuals in the city face an income reduction. As a result, your earnings ranking increases from 50% to 60%. After this event, you think you would be:

- ○ Better off
- ○ Slightly better off
- ○ The same
- ○ Slightly worse off
- ○ Worse off

Please indicate the degree to which you agree or disagree with the following statements:

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I like competition	○	○	○	○	○
I am a competitive individual	○	○	○	○	○
I enjoy competing against an opponent	○	○	○	○	○
I don't like competing against other people	○	○	○	○	○
I get satisfaction from competing with others	○	○	○	○	○
I find competitive situations unpleasant	○	○	○	○	○

>>

We are almost done, this is the last question of the survey. Please indicate the degree to which you agree or disagree with the following statements:

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I admire people who own expensive homes, cars, and clothes	○	○	○	○	○
The things I own say a lot about how I'm doing in life	○	○	○	○	○
Buying things gives me a lot of pleasure	○	○	○	○	○
I like a lot of luxury in my life	○	○	○	○	○
My life would be better if I owned certain things I don't have	○	○	○	○	○
I'd be happier if I could afford to buy more things	○	○	○	○	○

A.3 Sample Invitation Email to Deans

Dear Dean X,

I am a Graduate Student in Economics at the University of Illinois. Along with Ricardo Perez-Truglia (Economics Professor at University of California Los Angeles, Anderson School of Management), we are working on a research project about how people make important life decisions. I am writing you in your capacity as Dean in the hope that you would allow us to survey the students at University X about their choices in the National Residency Matching Program next January, before the ranking submission window opens.

Completing the online survey would take the students less than 10 minutes, and as a token of appreciation, we will send each respondent a $10 Amazon gift card. I have attached a draft of the survey for your reference. The questions are non-controversial, responses will remain strictly confidential, and we are open to incorporating your feedback into the survey.

The NRMP provides a perfect context to study important life decisions. We hope that the results from our study could provide useful information and insights to future generations of medical students applying to residency programs, and provide new insights to residency programs.

If you have any questions about the survey, we would be happy to answer them over e-mail or schedule a time for a brief phone conversation. We will be surveying students from medical schools around the country, and would love to add University X to our list of participating medical schools. Can we please count with your collaboration?

Best regards,

A.4 Sample Email to Students with Invitation to Baseline Survey

Dear graduating medical student,

We would like to invite you to participate in a brief, confidential survey about the Main Residency Match. **It takes less than 10 minutes to complete the survey and, as a token of our appreciation, respondents will be sent a $10 Amazon gift card by email.**

To participate in the survey, you must be registered in the 2017 Main Residency Match. If you want to participate, you must fill out the survey before you submit your rankings to the NRMP.

The survey can be accessed here: [LINK]

The results of this study will provide better information on how medical students select residency programs, and can assist in the advising and preparation of future generations of students.

We thank you and deeply appreciate your time and participation,

Ricardo Perez-Truglia, University of California, Los Angeles
Nicolas Bottan, University of Illinois at Urbana-Champaign
[Project's URL]

A.5 Sample Email to Students with Invitation to Follow-Up Survey

Dear graduating medical student,

Thank you for participating in our study! We wanted to invite you to participate in a very short follow-up survey. Your participation is voluntary and all responses will be kept strictly confidential. **It takes less than 5 minutes to complete the survey, and, as a token of our appreciation, we will send you a $5 Amazon gift card by email.**

Follow this link to the Survey: [LINK]

Or copy and paste the URL below into your internet browser: [URL]

After you complete this follow-up survey, your contact information will erased and we will not contact you again.

We thank you again and deeply appreciate your time and participation,

Ricardo Perez-Truglia, University of California, Los Angeles
Nicolas Bottan, University of Illinois at Urbana-Champaign
[Project's URL]
[Unsubscribe LINK]

A.6 Project's Website

UCLAAnderson
SCHOOL *of* MANAGEMENT

APPLY | FOR COMPANIES | GIVE

Global Economics and Management

- Overview
- › Faculty
- Courses
- University of California GEM-BPP Research Workshop
- › GEM Seminar
- Student Workshop
- Ph.D. Program
- Ph.D. Students
- Ph.D. Placements
- Working Papers
- FAQ
- Contacts

Details about the Residency Survey

We have been authorized by administrators in your medical program to invite you to participate in our survey that has the objective of better understanding how residency applicants form their NRMP rankings.

This study was approved by the UCLA Institutional Review Board. Your participation is voluntary, and is greatly appreciated: while it will not benefit you personally, it will help inform our research on the important process of deciding how to rank medical programs, which may benefit other medical students and medical programs in the future. You may withdraw from the study at any time.

Your privacy is very important to us. When information is transferred online there is a possibility that it may be viewed by a third party. To reduce the risk that an outside party could identify you or observe your responses, this survey employs Transport Layer Security (TLS) encryption for all transmitted data. As a result, we anticipate that your participation in this survey presents no greater risk than everyday use of the Internet. Your responses will be used solely for research purposes and will be kept strictly confidential, shared only with the researchers named below.

This study is being conducted by Ricardo Perez-Truglia (Assistant Professor at University of California, Los Angeles) and Nicolas Bottan (Ph.D. Candidate at the University of Illinois). If you have any questions or concerns about this survey, please contact us at: ricardo.truglia@anderson.ucla.edu or bottan2@illinois.edu.

If you have questions about your rights while taking part in this study, or you have concerns or suggestions and you want to talk to someone other than the researchers about the study, please call the OHRPP at (310) 825-7122 or write to: UCLA Office of the Human Research Protection Program, 11000 Kinross Avenue, Suite 211, Box 951694, Los Angeles, CA 90095-1694 (Ref: project 16-001968).

To be eligible to participate in this survey, you must be a medical student participating in the 2017 Main Residency Match and have not submitted your ranking order. Participants completing the entire survey will be paid $10 in the form of a Amazon Gift Card, that will be sent by email as soon as possible (Note: to receive payment, you must have a .edu email address).

Thank you for your attention,

Ricardo Perez-Truglia and Nicolas Bottan (The Research Team)

FOR VISITORS
campus tour
maps & directions
master calendar
facility use

FOR COMPANIES
recruit an mba
post a job
consulting teams
for GAP companies

FOR THE NEWS MEDIA
media relations
ucla anderson forecast
anderson in the news
faculty directory
faculty directory (pdf)
fact sheet

directory
site index
portal
library
UCLA
feedback
© UC Regents

B Information about the Subject Pool

We recruited 27 of the 135 accredited medical schools in the U.S. to participate in our study. In order to compare school characteristics from our sample with those not participating in our study, we obtained data from U.S. News (that is best known for compiling data and publishing ranks for universities and hospitals). We present descriptive statistics for the universe of medical schools, non-participating and participating schools in Table B.2. Medical schools participating in our study have slightly higher enrollment, lower average MCAT score, and are a little lower ranked on average than non-participating schools. However, none of these differences are statistically significant at conventional levels. The only statistically significant difference we do find is that the faculty to student ratio in participating schools is lower than in non-participating schools. Overall, it seems that participating medical schools are fairly representative of the overall universe of schools and not substantially different from non-participating schools.

Next, in Table B.1, we present the list of participating medical schools, along with the estimated size of the senior cohort, number of finished surveys and response rates. Around half of the schools reported the exact number of senior students who were participating in the Main Residency Match. For the remaining schools, we imputed the values of these variable using the average for the reporting schools (22% of the total enrollment). The overall response rate was almost 30%. Note that in the table we are excluding 20 observations that were deemed invalid either because answers to key questions were missing or feedback did not display correctly. These issues were due to technical difficulties most likely due to using a outdated internet browser without the proper Javascript support required to display and interact correctly with the survey. We have significant variation in response rates across medical schools. The response rate at Penn State is particularly low due to the fact that instead of forwarding the invitation by email, fliers were posted in the student lounge.

The day after the rank order submission deadline to the NRMP, we sent email invitations to the follow-up survey directly to respondents who had participated in the baseline survey. In Table B.3, we present descriptive statistics for our entire sample, and by whether respondents participated in the follow-up or not. The overall response rate to the follow-up was 90.6%. We do not find any statistically significant differences between the follow-up and non-follow-up respondents for all variables with the exception for single, where it appears that single students were less likely to participate in the follow-up survey. Additionally, participants to the follow-up survey reported slightly higher prior beliefs in cost of living than non-follow-up respondents. However, they were similarly "accurate" in their prior belief of cost of living.

Figure B.1: Distribution of Medical Schools in the U.S.

Notes: Each dot represents one of the 135 accredited medical schools contacted to participate in the study (excluding one in Hawaii). Dots do not denote exact location since they were moved to avoid overlap. Dark dots denote medical schools that agreed to participate in our study.

Table B.1: Survey Participation

State	University	Est. Senior Cohort	Nr Finished Surveys	Est. Response Rate (%)
Alabama	University of Alabama	174	47	27.0
Alabama	University of South Alabama	73	21	28.8
Arizona	University of Arizona	72	18	25.0
California	UC San Diego	124	39	31.5
Connecticut	Yale University	121	24	19.8
Florida	University of Florida	135	52	38.5
Illinois	Loyola University	145	66	45.5
Illinois	University of Illinois	20	8	40.0
Indiana	Indiana University	345	89	25.8
Massachusetts	Tufts University	194	42	21.6
Michigan	Michigan State University	183	76	41.5
Missouri	Saint Louis University	165	70	42.4
Missouri	University of Missouri (Kansas City)	101	34	33.7
Nebraska	University of Nebraska	125	46	36.8
New Mexico	University of New Mexico	97	27	27.8
New York	Stony Brook University	126	16	12.7
New York	University of Rochester	103	37	35.9
Ohio	Ohio State University	172	61	35.5
Oklahoma	University of Oklahoma	147	47	32.0
Pennsylvania	Pennsylvania State University	139	4	2.9
Rhode Island	Brown University	126	34	27.0
South Carolina	University of South Carolina	90	21	23.3
Texas	Baylor	180	44	24.4
Texas	Paul L. Foster School of Medicine (TTU)	89	30	33.7
Vermont	University of Vermont	105	39	37.1
Virginia	Virginia Commonwealth University	215	65	30.2
West Virginia	West Virginia University	110	23	20.9
	Total	3,676	1,080	29.38

Notes: 20 responses were excluded because they were deemed invalid (e.g., they did not received feedback due to a technical issue with their Internet Browser). Estimated senior cohort based on actual cohort size for schools that reported, and estimated as 22% of total enrollment for those that did not report cohort size (where 22% is the average proportion of seniors to total enrollment for schools that reported senior cohort size).

Table B.2: Comparison of Characteristics between Participating and Non-Participating Medical Schools

	All schools	Non-Participants	Participants	P-value
Enrollment	630.98	619.338	671.727	0.398
	(23.117)	(24.891)	(57.213)	
NR	0.267	0.287	0.185	0.245
	(0.038)	(0.044)	(0.076)	
Avg. MCAT	32.222	32.364	31.727	0.253
	(0.252)	(0.293)	(0.475)	
NR	0.267	0.287	0.185	0.245
	(0.038)	(0.044)	(0.076)	
Undergrad GPA	3.735	3.734	3.737	0.902
	(0.009)	(0.010)	(0.019)	
NR	0.267	0.287	0.185	0.245
	(0.038)	(0.044)	(0.076)	
Acceptance rate	0.066	0.067	0.062	0.458
	(0.003)	(0.004)	(0.005)	
NR	0.274	0.296	0.185	0.206
	(0.039)	(0.044)	(0.076)	
US News Ranking	45.451	43.478	51.636	0.166
	(2.784)	(3.309)	(4.872)	
NR	0.326	0.361	0.185	0.049
	(0.040)	(0.046)	(0.076)	
Tuition	51,404.98	51,333.526	51,651.818	0.913
	(1,097.842)	(1,193.139)	(2,689.180)	
NR	0.274	0.296	0.185	0.206
	(0.039)	(0.044)	(0.076)	
Faculty per student	2.363	2.518	1.827	0.039
	(0.221)	(0.279)	(0.177)	
NR	0.274	0.296	0.185	0.206
	(0.039)	(0.044)	(0.076)	
Peer Assessment score	3.14	3.139	3.145	0.961
	(0.076)	(0.093)	(0.106)	
NR	0.222	0.231	0.185	0.59
	(0.036)	(0.041)	(0.076)	
Observations	135	108	27	

Notes: Data for 135 accredited medical schools contacted by authors to participate in study. Data obtained from U.S. News for 2016. NR indicates the proportion of observations for which the statistic was either not published or missing. P-value in final column for the difference in means between participating and non-participating medical schools. Standard deviations reported in parenthesis.

Table B.3: Comparison of Characteristics between Respondents to Baseline and Follow-Up Surveys

	All	No Follow-up	Follow-up	P-value
Male (=1)	0.481	0.505	0.479	0.621
	(0.015)	(0.050)	(0.016)	
Age	27.091	26.921	27.108	0.482
	(0.083)	(0.253)	(0.088)	
Nr Kids	0.132	0.079	0.138	0.160
	(0.014)	(0.039)	(0.015)	
Single (=1)	0.354	0.505	0.338	0.001
	(0.015)	(0.050)	(0.015)	
Dual match (=1)	0.074	0.079	0.074	0.841
	(0.008)	(0.027)	(0.008)	
RPP treatment (=1)	0.499	0.525	0.496	0.588
	(0.015)	(0.050)	(0.016)	
ACS treatment (=1)	0.500	0.475	0.503	0.601
	(0.015)	(0.050)	(0.016)	
Average Residency Salary ($1000s)	0.013	0.019	0.012	0.871
	(0.013)	(0.042)	(0.014)	
Relative residency percentile	0.025	0.026	0.024	0.944
	(0.007)	(0.025)	(0.007)	
Pass Attention Check (=1)	0.964	0.950	0.965	0.509
	(0.006)	(0.022)	(0.006)	
Prior $ER_{1,2}$	0.004	0.008	0.004	0.775
	(0.005)	(0.013)	(0.005)	
Prior $COL_{1,2}$	0.004	-0.007	0.005	-0.479
	(0.006)	(0.016)	(0.007)	
Posterior $ER_{1,2}$	-0.009	-0.012	-0.008	0.639
	(0.003)	(0.008)	(0.003)	
Posterior $COL_{1,2}$	0.010	0.008	0.010	0.856
	(0.004)	(0.014)	(0.004)	
Observations	1,080	101	979	

Notes: Standard deviations reported in parenthesis. P-values correspond to the test of the null hypothesis of equal means between follow-up and non-follow-up samples. Relative residency percentile based on residency quality ranks computed by Doximity. All variables constructed with data from the baseline survey.

Table B.4: Comparison of Characteristics between Experimental Subjects and U.S. Population of Earners

	Survey	ACS 2015
	Med. Students	Adult Earners
Age	27.091	41.258
	(2.725)	(12.330)
% Male	0.481	0.515
	(0.500)	(0.500)
% Married	0.240	0.531
	(0.427)	(0.499)
Wage	54,203.4	50,877.0
	(3,447.0)	(56,438.8)
US Born	0.950	0.809
	(0.218)	(0.393)
% More than College	1	0.125
	(0.000)	(0.331)

Notes: Data from 2015 American Community Survey PUMS for the subsample of adults in between 21 and 65 years of age and who receive positive wage income.

C Estimation of the Feedback Provided to Subjects

C.1 Earnings Rank

To provide feedback on the earnings rank of each metropolitan and wage offered at the location, we used data for the American Community Survey (ACS) at the metro area level for 2015 and the latest data from the Current Population Survey (CPS), as stated in the debriefing message.[50] From the data we estimated the parameters (μ and σ) for fitting a log-normal distribution. In the ACS we based this on the proportion of total full-time year round workers with earnings in each earnings bin, over which we estimated the parameters of fitting a log-normal distribution using maximum-likelihood for each metropolitan area. For the CPS, we combined weekly earnings with overtime earnings in order to obtain as close a measure as possible to that in the ACS. We obtained the parameters for fitting a log-normal distribution by estimating, for each metro area, a right-censored Tobit of annualized log earnings on the intercept. In the ACS, only 2% of metro areas were missing, while 20% of metro areas were missing for the CPS. Most of the metro areas with missing values from the ACS were imputed using the corresponding values obtained from the 2011-2015 5-year ACS.[51] The missing values in the CPS were imputed using the state-averages.

From these parameters, given a wage in dollars, it is easy to calculate the percentile using a simple formula: $\Phi(log(wage_j) - \mu_j)/\sigma_j$. This way it was possible to provide personalized feedback according to different wages reported in the survey. For our sample of metro areas, the average percentile rank for earnings of \$55,000 is 59.2% and 68.9% for the ACS and CPS, and the correlation is 0.91.

Although both sources are similar in levels, there is plenty of exogenous variation between them when comparing pairwise differences of chosen locations. We show this variation in Figure C.1.a, where the R-squared of regressing the pairwise differences for the ACS on the pairwise differences for the CPS is 0.430.

C.2 Cost of Living

To provide feedback on cost of living in the metropolitan areas we use the Regional Price Parity Index (RPP) compiled by the Bureau of Economic Analysis and the Cost of Living Index (COLI). The Cost of Living Index has been published since 1968 (formerly known as ACCRA) and has been used extensively in academic research. For the Regional Price Parity Index we used their final index for 2014 (the latest available at the time we conducted the

[50] At the time, the latest two months available were September and October of 2016.

[51] Only 3 metro areas were still missing for which we imputed with the average values for the country.

survey), and for the Cost of Living Index we used their raw data for the first quarter of 2016, and calculated our own index by computing a weighted average over the expenditure categories grocery, utilities, transport, health, and miscellaneous (excluding housing).

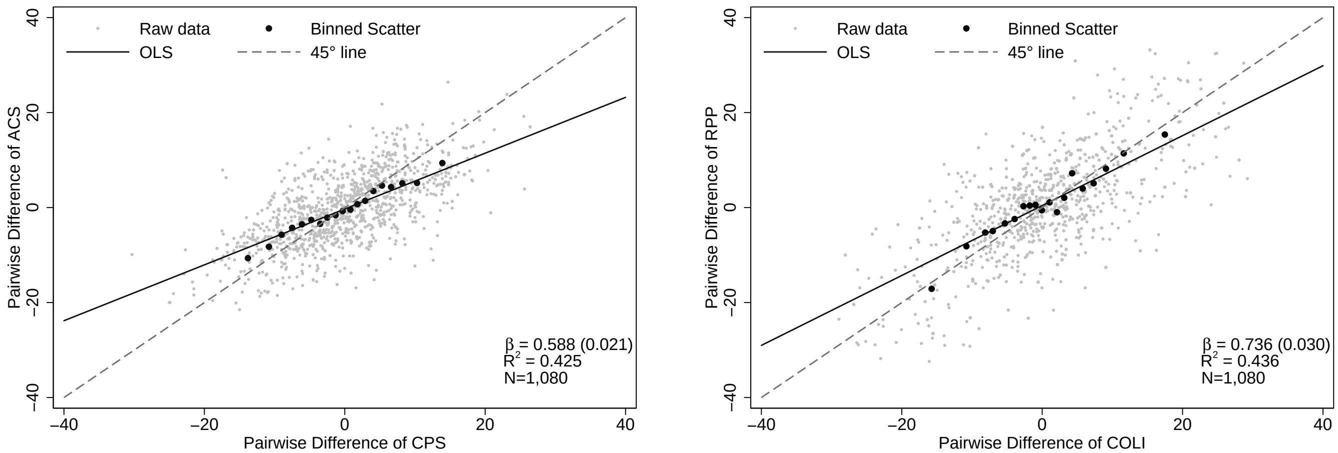
Both sources are quite similar: for the sample of potential metro areas that respondents can choose from based on the residency programs available, the mean cost of living for the RPP and COLI are 102.4 and 110.2, while the correlation between levels is 0.95. Note that the original indices have an average of 100. However, our sample of metro areas only include those for which there was a potential residency program to apply to. Therefore only 286 metro areas are included in our sample. These are mostly large and more expensive metro areas, which explains why the average is larger than 100. Additionally, 37 and 117 metro areas were imputed for RPP and COLI, respectively. We imputed values using predictions based on OLS regressions that included metro area census characteristics such as population, average household size, income, population density, racial and educational composition, housing characteristics and state dummies. The R^2 for those regressions was 93% and 86%. In our survey, less than 1% of metro options receiving RPP feedback were imputed, while only 11% of COLI feedback metro options were imputed.

Even though both sources are similar when comparing them in levels, there is substantial exogenous variation when comparing the pairwise differences between chosen locations. The variation is presented in Figure C.1.b, where the R-squared for regressing the pairwise differences of RPP on COLI is only 0.436.

Figure C.1: Comparison of Statistics between the Two Different Data Sources

a. Earnings Rank

b. Cost of Living



Notes: Pairwise differences of statistics from different sources of cost of living and earnings rank based on cities and wages indicated by respondents in the survey. The gray dots correspond to the raw scatterplot, and the darker dots correspond to the binned-scatterplot based on 20 bins. The sources in Panel a correspond to ACS (American Community Survey) and CPS (Current Population Survey). The sources from Panel b correspond to RPP (Regional Price Parity Index) and COLI (Cost of Living Index). The slope (β, with robust standard errors in parentheses) and R^2 are based on a linear regression.

D Additional Results

D.1 Variation in Nominal Income, Cost of Living and Earnings Rank

Residency programs do not compensate for differences in cost of living or earnings rank through wages. Figure D.1.a presents a scatterplot of the the costs of living versus the (log) nominal residency income. The strength of the association between these two variables represent the degree to which residencies compensate for local cost of living. The low value of the R^2 (0.115) shows that residencies compensate only partially, leaving substantial orthogonal variation between cost of living differences and differences in nominal income.

In a similar spirit, Figure D.1.b explores the extent to which programs compensate for differences in the distribution of income through their nominal wages. This figure shows a scatterplot of the earnings rank at the residency income versus the nominal income. The strength of the association indicates how much of the differences in earnings rank are generated by differences in nominal income. Again, the low value of the R^2 (0.011) indicates that the vast majority of the variation in earnings rank is orthogonal to the nominal income.

D.2 Accuracy of Prior Beliefs, Pairwise Differences

Respondents may have a poor idea of the levels of cost of living and earnings rank, but they may have a better understanding of relative differences—ultimately the relevant statistic in decision making. We repeat our previous analysis, examining the pairwise differences instead of levels. in Figure D.2.a for cost of living and Figure D.2.b for earnings rank. Although the results for cost of living are almost identical, prior beliefs about earnings ranks are somewhat more accurate with pairwise differences. For example, the slope coefficient increases to 0.793, while the R^2 increases to 0.15 (5 times larger than in levels). In any case, even under this alternative specification, the accuracy of prior beliefs about earnings rank remain far less accurate than those for cost of living.

D.3 Learning Rates are Similar by Information Source

One concern with our experimental design is that individuals may have updated their beliefs differentially depending on the source used. For example, if respondents believe one source to be less trustworthy than another they may disregard that feedback. We explore the extent to which this could have happened by separately examining learning by the information source used. In Figures D.3 we present the same figures as in section 5.3 by information source.

Learning rates for cost of living and earnings ranking are almost identical between sources, showing that respondent's reactions to information did not depend on the source.

D.4 Beliefs are Persistent in Follow-up Survey

Since posterior beliefs on cost of living and earnings ranking were elicited directly after providing respondents feedback, we are interested in examining how persistent these beliefs are a month later. We show that posterior beliefs are persistent for both cost of living and earnings rank in Figure D.4. The persistence in cost of living is twice as large as that of earnings rank (correlation of 0.844 versus 0.464), most likely due to respondents reverting to their prior beliefs over time and the fact that their priors were significantly more accurate for cost of living than earnings rank.

D.5 Complementary Evidence: Hypothetical Questions

To provide some additional suggestive evidence that individuals care about their consumption rank, we included a couple of hypothetical questions at the end of the follow-up survey (for the exact wording of this question, see Questionnaire Appendix A.2).

In the first scenario, we elicited the subjects' preferences for a reduction in cost of living while holding the earnings rank constant. More precisely, we asked the respondents whether they would be better off, the same, or worse off if their own cost of living and the cost of living of all other individuals in the city went down by 10%. Figure D.5.a shows the distribution of responses. Consistent with preferences for absolute consumption, 80% of respondents answered that they would be better off with this change, with 19% reporting that they would be the same and less than 1% responding that they would be worse off.

In the second scenario, we elicited the subjects' preferences for an increase in earnings rank, while holding the cost of living constant. To do so, we asked the respondents whether they would be better off, the same, or worse off if their own income and cost of living stayed the same but all other individuals in the city faced an income reduction of 10%. Because of the social desirability bias, individuals may not want to "confess" so directly that they care about relative income, and thus these responses probably lead to an underestimation of concerns for relative income. Figure D.5.b shows the distribution of responses. Consistent with individuals having direct preferences over relative income, 44% of individuals responded that they would be either better or worse off, with significant heterogeneity. While 31% of individuals reported that they would be better off with the poorer neighbors, 13% of individuals reported that they would be worse off.

D.6 Marginal Effects

Given that it is not possible to directly interpret coefficients from Probit regressions, in Table D.1 we present estimates from columns (1) to (3) from Table 2 along with their corresponding marginal effects at the average in the first two rows. The third and fourth row restrict the sample to respondents from the follow-up survey. For example, the coefficient in column (3) for the baseline sample implies that an increase of 1 percentage point in earnings rank in location 1 would increase the probability of choosing that location by 0.186 percent (or, in other words, an implied behavioral elasticity of 0.186).

D.7 Preferences over Subjective Program Characteristics

To better understand the magnitude of our results, we compare the estimates for preferences over earnings rank and cost of living with those of subjective perceptions of residency program characteristics (prestige, career prospects and sense of purpose). These perceptions were elicited by the end of the follow-up survey and are standardized to have mean zero and standard deviation of one. We estimate the baseline model presented in section 3.1, introducing the three perceived program characteristics one by one. Since we only observe these perceptions in the follow-up survey, we restrict the sample to those respondents.

The results are presented in Table D.2. The coefficients on the three subjective perceptions of the program are positive, as expected, and highly statistically significant (all p-values$<$0.001). This means that individuals prefer programs associated with higher purpose, career prospects and prestige. Furthermore, we can compare the strength of these preferences to the strength of preferences for cost of living.[52]

We cannot compare the raw Probit coefficients directly, because the independent variables are measured in different units. For a meaningful comparison, we can calculate the standardized coefficient corresponding to a one standard deviation decrease in cost of living. According to column (2) of Table D.2, a one standard deviation decrease in cost of living would correspond to a Probit coefficient of 0.167 (i.e., the non-standardized coefficient, 1.211, multiplied by the standard deviation of cost of living, 0.138). This standardized coefficient can be compared to the coefficient of 0.441 corresponding to a one standard deviation increase in the sense of purpose. This comparison implies that the sense of purpose of a program is 2.64 times as important as the cost of living. By the same metric, the career prospects (column (3)) and sense of prestige (column (4)) are 2.27 and 1.49 times as important as cost of living. In sum, the characteristics of a program are systematically more important for the choice of residency than the cost of living during the residency.

[52] The results are similar if we do the comparison with respect to the preferences for earnings rank instead.

D.8 No Other Significant Preference Heterogeneity

In this section we explore additional heterogeneity over preferences for earnings rank and cost of living. We first decompose the results of heterogeneity by relationship status in two ways. In columns (1) and (2) of Table D.3, we show that within non-single respondents, preferences over relative income are similar for married or long-term relationship respondents. However, it seems that preferences for cost of living are mostly driven by married respondents (though the difference is borderline insignificant, p-value=0.109). In columns (3) to (6) of Table D.3, we estimate preferences by gender, within relationship status. Preferences over earnings rank seem to be stronger for females in general, though the difference is not statistically significant for non-singles or singles.

In addition to the dimensions explored in the paper, we present results for heterogeneity across different dimensions in Table D.4. In columns (1) to (4) we explore heterogeneity according to differences in hypothetical choices of changes in earnings rank and cost of living. Interestingly, we find that those who believe they would be better off if cost of living were to decrease care significantly more about earnings rank than respondents who claimed they would be the same or worse off. However, we do not find any significant differences for the hypothetical question of a change in earnings rank.

Next, we explore whether there is preference heterogeneity across different individual traits, such as degree of materialism, competitiveness or life dimensions valued the most. The materialism index is based on questions that typically reflect status from consumption (see follow-up survey questionnaire in Appendix A.2, based on Richins and Dawson, 1992). Even though we do not find statistically significant difference in the effects in columns (5) and (6), the point estimates are different and reflect that those who are classified as more "materialistic" (or in other words, those most concerned by the signaling value of material goods) care more about earnings rank, while those who are less "materialistic" care more about cost of living. In columns (7) and (8) we explore heterogeneity by the degree of competitiveness using commonly used indices in psychology (Smither and Houston, 1992). We do not find any significant differences across these traits. Finally, in columns (9) and (10) we explore heterogeneity according to a principal component score of the rank of different life dimensions by importance (happiness, health, sense of purpose, spirituality, control over life). We do not find any statistically significant differences in these dimensions.

D.9 Results are Robust to Dropping Specific Subgroups

In this section we explore the sensitivity of our baseline results to dropping specific subgroups that may potentially attenuate our estimates for preferences over earnings rank and cost of living. In the first row of Table D.5 we report the baseline estimates. In the second row, we re-estimate the model dropping respondents that did not successfully answer a question at the end of the baseline survey designed to test whether they were paying attention and reading the questions carefully. In this question we describe how emotions can play a role in influencing responses and respondents have a menu of emotions to choose from. However, at the end of the paragraph we instruct respondents to only select the option "none of the above" (see Appendix A.1 for the full question). Only 3.6% of respondents failed to answer this question correctly. Estimates do not change much when dropping these respondents – if anything, the coefficients are slightly larger in magnitude.

One additional concern is that respondents may not choose according to their own preferences but define it jointly with their spouse when they are both participating as a dual match. In the third row of Table D.5 we drop respondents who are participating in a dual match (7.4% of the sample). Again, the results are similar when we drop these respondents.

D.10 Results are Similar when using Binary Probit or Ordered Probit

In the baseline survey we asked respondents about their intention to rank using a likelihood scale, that we later converted in to a binary variable in order to directly compare it to their final rank submission in the follow-up survey. However, we could also exploit the full variation of using the likelihood scale by means of estimating an ordered Probit model. The results are presented in Table D.6. Overall, the results are quite similar regardless of using the binary or likelihood variables.

D.11 Instrumental Variable Regression

We break down the Instrumental Variables regression into the first-stage and reduced-form regressions. Table D.7.a presents the same experimental estimates as those found in the second row of Table 4. In the next panel we focus on the first stages. As discussed in section 5.3, respondents learn from our information provision experiment, where learning rates are close to 1 for both earnings rank and cost of living. It does not seem that weak instruments are a problem overall. However, the instruments are substantially weaker for the sample of singles compared to the non-singles, where the Cragg-Donald F-statistic drops from 169 to 42. In the final panel of Table 4 we show that the reduced form estimates are similar to those obtained by IV.

D.12 Comparison to Studies using Subjective Data

We are interested in comparing our results to those obtain in previous studies based on happiness surveys or hypothetical choices. It is important to note that these other studies measure relative concerns in a slightly different way. They compare the effects of own consumption versus the mean consumption of peers. They present an econometric model along the following lines:

$$U = a \cdot log(y) - b \cdot log(\bar{y})$$

Where y is the individual's own income and $\bar{y}$ is the average income in the individual's reference group. With parameters a and b, we can calculate the trade-off between absolute and relative income. The effect of absolute income is given by $a-b$: i.e., what would happen if increase my income by 1% if I am also increasing everyone else's income by 1%. The effect of relative income is just b: i.e., what happens if you increase everyone else's income by 1% while leaving my own income unchanged. An individual with parameters a and b should be indifferent between a 1% increase in her absolute consumption and a $\frac{a-b}{b}$ decrease in her relative consumption. Table D.8 shows the estimates of a and b reported in other studies, and the resulting estimate of $\frac{a-b}{b}$.[53]

Section 8.2 compares our estimates with respect to the findings from Luttmer (2005). In this section, we provide comparisons with respect to other studies. According to our baseline estimates for non-singles (column (1) of Table 4), the average individual is willing to give up 1 percent of her absolute consumption to decrease the median consumption of her peers by 0.91%.[54] The other studies that use happiness data suggest a corresponding trade-off of 0.89% (Clark, Senik and Yamada, 2017) and 1.02% (Ferrer-i-Carbonell, 2005); while the studies using hypothetical choices suggest a corresponding trade-off of 1.85% (Johansson-Stenman et al., 2002) and 1.18% (Yamada and Sato, 2013). All of these estimates are in the ballpark of our own estimate of 0.91%, implying that, relative to these other studies, our estimates suggest a similar role for relative concerns.

Last, we must note that some studies find the opposite effect. For instance, Senik (2004) and Clark, Kristensen and Westergård-Nielsen (2009) find that life satisfaction is increasing in the mean income of the reference group. And Shigeoka and Yamada (2016) show estimates from a hypothetical choice experiment with mixed results: while the U.K. respondents prefer

[53]The table does not include standard errors or confidence intervals because we do not have sufficient information to compute those ($\frac{a-b}{b}$ is a non-linear function, and thus it does not suffice with the standard errors of a and b).

[54]This result arises because, for the average individual in the sample, we would need to decrease the median earnings in the area by 0.91% to allow the individual to climb up 0.518 (= 1/1.928) percentage points in the earnings rank.

poorer peers, the opposite is true for their U.S. respondents.

D.13 Estimated Preferences are similar when using Choice or Happiness

We can also exploit a different outcome variable, the happiness rank between the options, to compare the preferences inferred from choice versus happiness. Consistent with Benjamin et al. (2014), we observe a significant correlation (0.456) between the choice ranks and happiness ranks of these individuals. However, this association is far from perfect, which suggest that individuals are not choosing to maximize their happiness only. As a result, it is not obvious that preferences inferred from choice will be similar to preferences inferred from happiness.

Table D.9 presents results using happiness as outcome variables. These coefficients are of course not directly comparable to those of choice, because they are based on different dependent variables with different distributions. The baseline preferences are roughly consistent. For instance, for the full sample, β^{ER} is 1.141 (s.e. 0.577) for choice and 0.936 (s.e. 0.520) for happiness; while β^{COL} is -1.262 (s.e. 0.531) for choice and -0.311 (s.e. 0.479) for happiness. We cannot reject the null hypotheses that these two pairs of coefficients are equal. This evidence suggests that the happiness and choice trade-offs may be similar – however, given the precision of the estimates, we cannot reject the possibility of substantial discrepancies.

D.14 Recruitment of Auxiliary Experiment

We conducted an auxiliary experiment using a sample of respondents from Amazon Mechanical Turk (mTurk), an online job market for crowdsourcing small tasks. During September of 2017, we recruited the auxiliary sample through work postings (or HIT - “Human Intelligence Task”) on mTurk. Participants were invited to participate in a 8 minute survey “about city perceptions”. When accepting the task, participants were re-directed to the survey. After successful completion of the survey, participants were given a code to redeem their payment of $0.60 for completing the task. We restricted the survey to participants located in the United States.

D.15 Sample Questionnaire: Auxiliary Experiment

To get a general picture of the people answering this survey, we would like to ask you a few things about yourself. Please remember that your answers are confidential and that your name is not collected as part of this study.

Please indicate your gender:

- ○ Female
- ○ Male

How old are you?

Age

How many children do you have? is below; first:

What is your relationship status?

- ○ Single
- ○ In a long-term relationship
- ○ Married

How many children do you have?

How many children do you have?

Recent research on decision making shows that choices are affected by the context in which they are made. Differences in how people feel, in their previous knowledge and experience, and in their environment can influence the choices they make. To help us understand how people make decisions, we are interested in information about you, specifically whether you actually take the time to read the instructions; if you don't, some results may fail to tell us very much about decision making in the real world. To help us confirm that you have read these instructions, please ignore the question below about how you are feeling and instead check only the "none of the above" option. Thank you very much.

- ☐ Interested
- ☐ Enthusiastic
- ☐ Inspired
- ☐ Distressed
- ☐ Proud
- ☐ Determined
- ☐ Excited
- ☐ Irritable
- ☐ Attentive
- ☐ Scared
- ☐ Alert
- ☐ None of the above

>>

Consider the following hypothetical scenario:

Think about two cities you know well but that you do not currently live in.

Now imagine that you are offered a job where you will be paid **an annual gross salary of $54,000** and have to move to one of those two cities.

Please tell us (in no particular order) the first location you would consider moving to:

State: Illinois

Metropolitan Area: Champaign-Urbana, IL

How well do you know this place? [0 - not at all; 10 - Extremely well]

0 1 2 3 4 5 6 7 8 9 10

>>

Please tell us the other location you would consider moving to:

State: California

Metropolitan Area: Los Angeles-Long Beach-Anaheim, CA

How well do you know this place? [0 - not at all; 10 - Extremely well]

0 1 2 3 4 5 6 7 8 9 10

>>

Now, we want to ask you a couple of questions about the two cities you are considering living in.

Let's start with the expected cost of living. You probably noticed that the average prices of goods and services are different across different cities. As a result, with the same income, you would be able to buy more things in some cities and less in other cities.

Imagine that you chose to work in the **Champaign-Urbana, IL** metro area. Would you expect your cost of living in this city to be cheaper or more expensive than the U.S. average?

- ● cheaper
- ○ more expensive

How much cheaper is the Champaign-Urbana, IL metro area than the U.S. average?

\>\>

5

Imagine that you chose to work in the **Los Angeles-Long Beach-Anaheim, CA** metro area. Would you expect your cost of living in this city to be cheaper or more expensive than the U.S. average?

- ○ cheaper
- ● more expensive

How much more expensive is the Los Angeles-Long Beach-Anaheim, CA metro area than the U.S. average?

>>

Now we want to ask you about your expected earnings ranking. This ranking is defined as the share of the working individuals of a city who earn less than you. You probably noticed that the distribution of earnings is different across different cities. As a result, with the same earnings, you may be relatively rich in some cities but relatively poor in other cities.

Imagine that you chose to work in **Champaign-Urbana, IL**. With your individual annual earnings of **$ 54,000**, you would be richer than what percentage of **Champaign-Urbana, IL**'s individual earners?

Richer than 70% of individual earners

>>

Imagine that you chose to work in **Los Angeles-Long Beach-Anaheim, CA**. With your individual annual earnings of **$ 54,000**, you would be richer than what percentage of **Los Angeles-Long Beach-Anaheim, CA**'s individual earners?

Richer than 45% of individual earners

>>

Now, we want to share some information with you, related to the characteristics of the two cities that you are considering living in. Please take a moment to review the information carefully.

Note*: this information is only shown once and you will not be able to come back to it.*

First, find below some estimates of the cost of living:

The **Champaign-Urbana, IL** metro area is **6.6% cheaper** than the U.S. average.

The **Los Angeles-Long Beach-Anaheim, CA** metro area is **17.0% more expensive** than the U.S. average.

Source: based on most recent data from the Bureau of Economic Analysis.

Second, find below some estimates of the earnings ranking:

With your individual annual earnings of **$ 54,000**, you would be richer than **60.3%** of **Champaign-Urbana, IL**'s population.

With your individual annual earnings of **$ 54,000**, you would be richer than **57.9%** of **Los Angeles-Long Beach-Anaheim, CA**'s population.

Source: based on most recent data from the American Community Survey.

That was all the information that we wanted to share with you. Now that you have reviewed this information, we would like to ask you again about your expected cost of living and earning rankings.

Let's start with the cost of living:

Imagine that you chose to work in the **Champaign-Urbana, IL** metro area. Would you expect your cost of living in this city to be cheaper or more expensive than the U.S. average?

- ● cheaper
- ○ more expensive

How much cheaper is the Champaign-Urbana, IL metro area than the U.S. average?

6%

>>

Imagine that you chose to work in the **Los Angeles-Long Beach-Anaheim, CA** metro area. Would you expect your cost of living in this city to be cheaper or more expensive than the U.S. average?

- ○ cheaper
- ● more expensive

How much more expensive is the Los Angeles-Long Beach-Anaheim, CA metro area than the U.S. average?

>>

Imagine that you chose to work in **Los Angeles-Long Beach-Anaheim, CA**. With your individual annual earnings of **$ 54,000**, you would be richer than what percentage of **Los Angeles-Long Beach-Anaheim, CA**'s individual earners?

>>

Imagine that you chose to work in **Champaign-Urbana, IL**. With your individual annual earnings of **$ 54,000**, you would be richer than what percentage of **Champaign-Urbana, IL**'s individual earners?

>>

12

We understand this is a lot of information to process, so we will help you make the comparison simpler. According to your final answers about incomes, cost of living and relative earnings:

- If you chose to live in Champaign-Urbana, IL, you would be able to afford 19.8% more than if you chose to live in Los Angeles-Long Beach-Anaheim, CA.

- If you chose to live in Champaign-Urbana, IL, your earnings ranking would be 9.3% higher than if you chose to live in Los Angeles-Long Beach-Anaheim, CA.

If given the choice, where would you choose to live?

- Very likely Champaign-Urbana, IL
- Likely Champaign-Urbana, IL
- Leaning Champaign-Urbana, IL
- Leaning Los Angeles-Long Beach-Anaheim, CA
- Likely Los Angeles-Long Beach-Anaheim, CA
- Very likely Los Angeles-Long Beach-Anaheim, CA

>>

Next, we want to ask you about your perceptions regarding other aspects of these two cities .

>>

Imagine that you chose to work in the **Champaign-Urbana, IL** metro area. Do you think that **healthcare** (such as quantity and quality of hospitals and doctors) is better or worse than the US average?

- ○ better
- ● worse

How much worse is healthcare than the U.S. average?

>>

Imagine that you chose to work in the **Los Angeles-Long Beach-Anaheim, CA** metro area. Do you think that **healthcare** (such as quantity and quality of hospitals and doctors) is better or worse than the US average?

- ● better
- ○ worse

How much better is healthcare than the U.S. average?

>>

15

Imagine that you chose to work in the **Champaign-Urbana, IL** metro area. Do you think that **schools** would be better or worse than the US average?

- [x] better
- [ ] worse

How much better are schools than the U.S. average?

>>

Imagine that you chose to work in the **Los Angeles-Long Beach-Anaheim, CA** metro area. Do you think that **schools** would be better or worse than the US average?

- [ ] better
- [x] worse

How much worse are schools than the U.S. average?

>>

Imagine that you chose to work in the **Champaign-Urbana, IL** metro area. Do you think that the **quality of the environmen**t (such as the air and water purity) is better or worse than the US average?

- [x] better
- [ ] worse

How much better is the quality of the environment than the U.S. average?

>>

Imagine that you chose to work in the **Los Angeles-Long Beach-Anaheim, CA** metro area. Do you think that the **quality of the environment** (such as the air and water purity) is better or worse than the US average?

- [x] better
- [ ] worse

How much better is the quality of the environment than the U.S. average?

>>

Imagine that you chose to work in the **Champaign-Urbana, IL** metro area. Do you think that **public spaces** (such as the number and quality of parks and recreational areas) are better or worse than the US average?

- ○ better
- ● worse

How much worse are public spaces than the U.S. average?

>>

Imagine that you chose to work in the **Los Angeles-Long Beach-Anaheim, CA** metro area. Do you think that **public spaces** (such as the number and quality of parks and recreational areas) are better or worse than the US average?

- ● better
- ○ worse

How much better are public spaces than the U.S. average?

>>

18

Imagine that you chose to work in the **Champaign-Urbana, IL** metro area. Do you think that the **share of population with a college degree** is higher or lower than the US average?

- ● higher
- ○ lower

How much higher than the U.S. average?

\>>

Imagine that you chose to work in the **Los Angeles-Long Beach-Anaheim, CA** metro area. Do you think that the **share of population with a college degree** is higher or lower than the US average?

- ○ higher
- ● lower

How much lower than the U.S. average?

\>>

Imagine that you chose to work in the **Champaign-Urbana, IL** metro area. Do you think that the **quality of entertainment** (such as the number and quality of cinemas, theaters and bars) is better or worse than the US average?

- ○ better
- ● worse

How much worse is the quality of entertainment than the U.S. average?

\>>

Imagine that you chose to work in the **Los Angeles-Long Beach-Anaheim, CA** metro area. Do you think that the **quality of entertainment** (such as the number and quality of cinemas, theaters and bars) is better or worse than the US average?

- ● better
- ○ worse

How much better is the quality of entertainment than the U.S. average?

\>>

Imagine that you chose to work in the **Champaign-Urbana, IL** metro area. Do you think that the **share of population who voted for Donald Trump in the 2016 general election** is higher or lower than the US average?

- ○ higher
- ● lower

How much lower than the U.S. average?

>>

Imagine that you chose to work in the **Los Angeles-Long Beach-Anaheim, CA** metro area. Do you think that the **share of population who voted for Donald Trump in the 2016 general election** is higher or lower than the US average?

- ○ higher
- ● lower

How much lower than the U.S. average?

>>

Imagine that you chose to work in the **Champaign-Urbana, IL** metro area. Do you think that **crime** is higher or lower than the US average?

- ○ higher
- ● lower

How much lower is crime than the U.S. average?

\>>

Imagine that you chose to work in the **Los Angeles-Long Beach-Anaheim, CA** metro area. Do you think that **crime** higher or lower than the US average?

- ● higher
- ○ lower

How much higher is crime than the U.S. average?

\>>

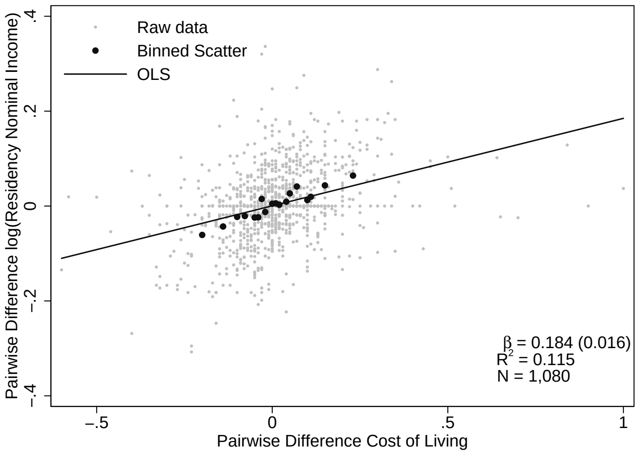
Thank you for participating.

Your validation code is:
[redacted]

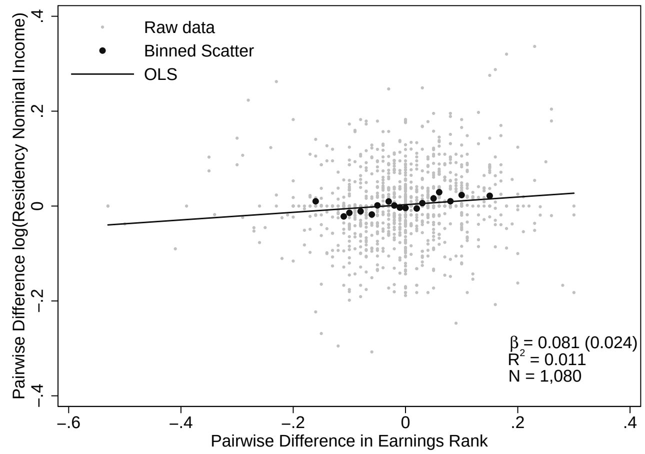
To receive payment for participating, click “Accept HIT” in the Mechanical Turk window, enter this validation code, then click “Submit”.

Figure D.1: Variation in Nominal Income, Cost of Living and Earnings Rank

a. Cost of Living vs. Nominal Income



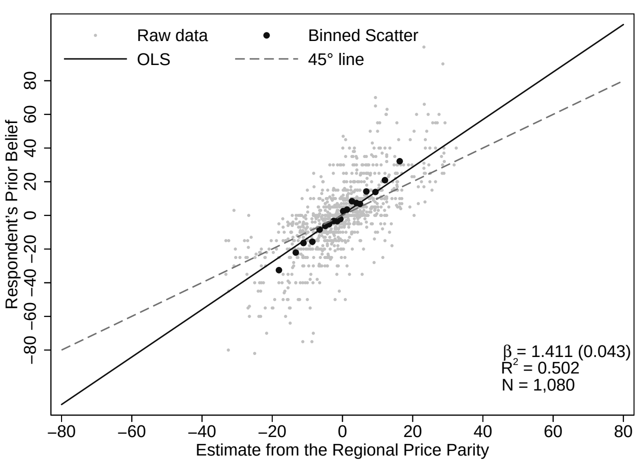
b. Earnings Rank vs. Nominal Income



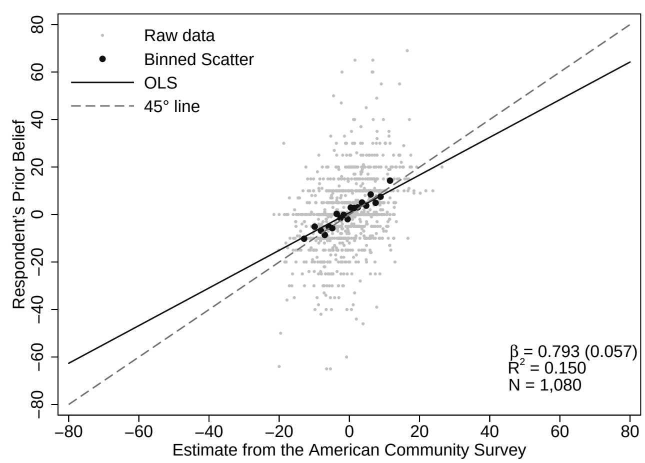
Notes: The gray dots correspond to the raw scatterplot, and the darker dots correspond to the binned-scatterplot based on 20 bins. Slopes (β, with robust standard errors in parentheses) and R^2 are based on a linear regression. All variables for x-axis and y-axis correspond to pairwise differences across the two cities that the subject is considering submitting to the algorithm. Data from survey responses, the Regional Price Parity Index (for cost of living) and the American Community Survey (for earnings rank).

Figure D.2: Comparison Between Prior Beliefs and Statistics

a. Cost of Living, Pairwise Differences

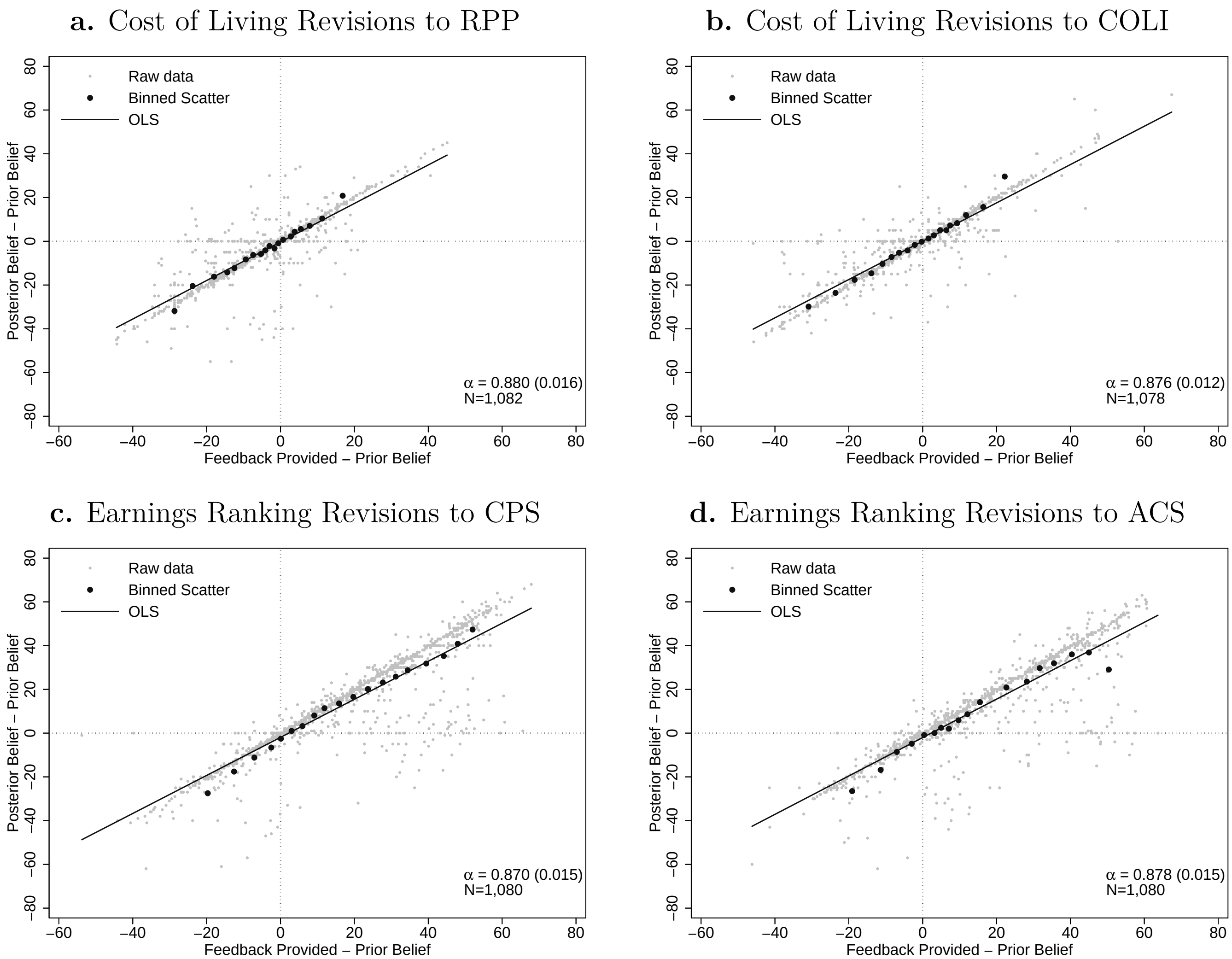


b. Earnings Rank, Pairwise Differences



Notes: Comparison between respondent's perceptions before the information provision (i.e., prior beliefs) and statistics. The gray dots correspond to the raw scatterplot, and the darker dots correspond to the binned-scatterplot based on 20 bins. Panels a and b present pairwise differences between an individual's options (i.e., value for first option minus that of the second option). The slope (β, with robust standard errors in parentheses) and R^2 are based on a linear regression.

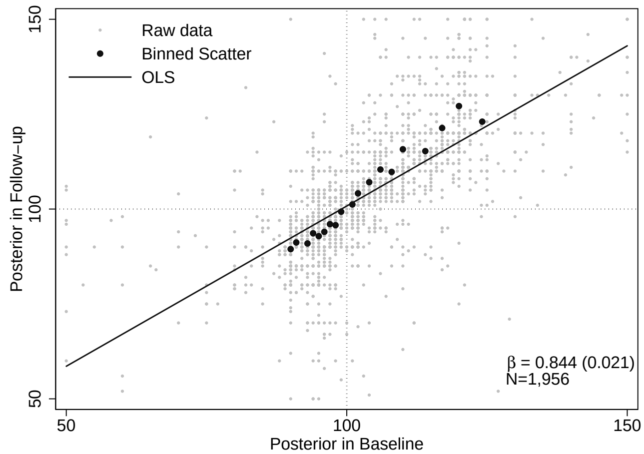
Figure D.3: Reduced-Form Evidence of Learning in the Information-Provision Experiment by Feedback Source



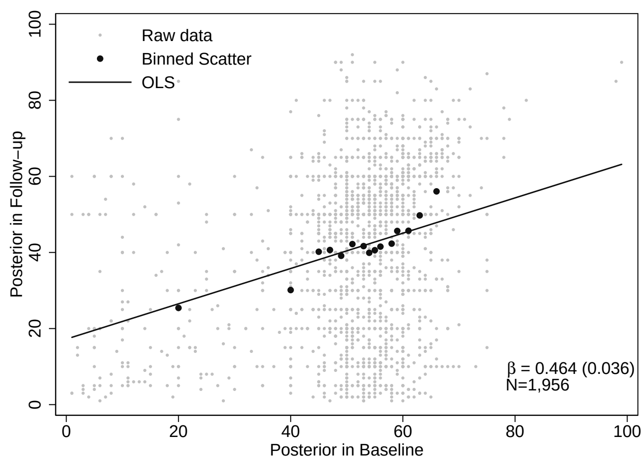
Notes: Comparison between the difference in statistics and respondent's perceptions before the information provision (i.e., prior beliefs), and difference in respondent's perceptions after the information provision (i.e., posterior beliefs) and prior beliefs. The gray dots correspond to the raw scatterplot, and the darker dots correspond to the binned-scatterplot based on 20 bins. Panels a and b show cost of living revisions to statistics from RPP (Regional Price Parity Index) and COLI (Cost of Living Index). Panels c and d show earnings rank revisions to statistics from CPS (Current Population Survey) and ACS (American Community Survey). The slope (α, with robust standard errors in parentheses) is based on a linear regression.

Figure D.4: Correlation between (Posterior) Beliefs in Baseline and Follow-Up Surveys

a. Cost of Living



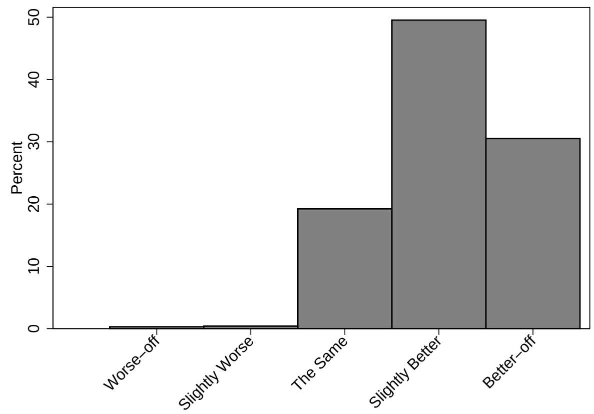
b. Earnings Rank



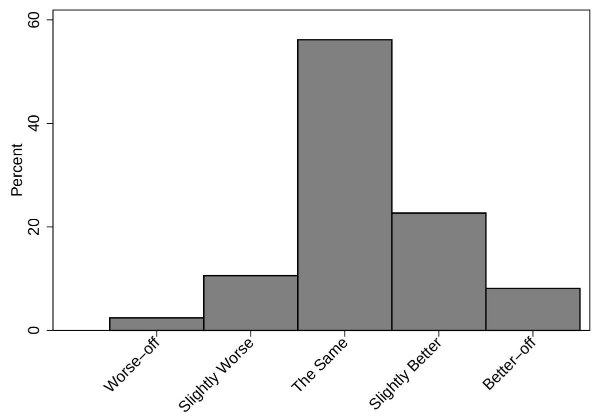
Notes: The gray dots correspond to the raw scatterplot, and the darker dots correspond to the binned-scatterplot based on 20 bins. Panels a and b present data in levels (i.e., two observations per individual, one for each of their options). The slope (β, with robust standard errors in parentheses) and R^2 are based on a linear regression.

Figure D.5: Survey Responses, Preferences over Hypothetical Changes to Cost of Living and Earnings Rank

a. Hypothetical Decrease in Cost of Living



b. Hypothetical Increase in Earnings Rank



Notes: Distribution of responses to hypothetical choice questions included in follow-up survey. Panel b corresponds to the question labeled "Event A", while panel b corresponds to the question labeled "Event B" in the questionnaire to the follow-up survey in Appendix A.2.

Table D.1: Probit Marginal Effects

	Panel A: β^{ER}			Panel B: β^{COL}		
	Non-Single (1)	Single (2)	All (3)	Non-Single (4)	Single (5)	All (6)
Baseline Sample						
Raw Probit	2.236***	-1.538*	0.995*	-1.087	-1.058	-1.073**
	(0.669)	(0.880)	(0.539)	(0.663)	(0.749)	(0.485)
Marginal Effect	0.418***	-0.267*	0.186*	-0.203	-0.183	-0.201**
	(0.125)	(0.155)	(0.100)	(0.124)	(0.130)	(0.090)
Follow-up Sample						
Raw Probit	2.380***	-1.656*	1.141**	-1.234*	-1.379*	-1.262**
	(0.702)	(0.991)	(0.577)	(0.743)	(0.772)	(0.531)
Marginal Effect	0.425***	-0.253*	0.202**	-0.221*	-0.211*	-0.224**
	(0.125)	(0.154)	(0.102)	(0.132)	(0.118)	(0.094)

Notes: Heteroskedasticity-robust standard errors in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Raw Probit coefficients and corresponding marginal effects at the mean. Probit regressions of expected rank order submission on posterior beliefs about cost of living and earnings rank estimated by sample (i.e., coefficients of a same row and sample are from a single regression). All specifications include the baseline controls listed in section 3. Results for Baseline Sample are based on the sample of individuals who completed the baseline survey (1,080 responses, 698 from non-singles and 382 from singles). Results for Follow-up Sample are based on the sample of individuals who completed the follow-up survey (978 responses, 647 from non-singles and 311 from singles).

Table D.2: Preferences for Subjective Program Characteristics

	(1)	(2)	(3)	(4)
β^{ER}	1.141**	1.147*	1.172*	1.172**
	(0.577)	(0.609)	(0.602)	(0.584)
β^{COL}	-1.262**	-1.211**	-1.470***	-1.412***
	(0.531)	(0.529)	(0.525)	(0.515)
$\beta^{purpose}$		0.441***		
		(0.064)		
$\beta^{prospects}$			0.379***	
			(0.070)	
$\beta^{prestige}$				0.249***
				(0.061)
Pseudo R^2	0.035	0.120	0.093	0.059
Observations	978	978	978	978

Notes: Heteroskedasticity-robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Raw Probit coefficients. Probit regressions of expected rank order submission on posterior beliefs about cost of living and earnings rank estimated by sample (i.e., coefficients of a same row and sample are from a single regression). All estimates, include the baseline controls listed in section 3. Mean (standard deviation) for $ER_{1,2}^{i,posterior}$ is -0.008 (0.098) and for $COL_{1,2}^{i,posterior}$ is 0.010 (0.138). Measures for subjective program characteristics (prestige, prospects, purpose) are standardized to have mean zero and standard deviation of one.

Table D.3: Preference Heterogeneity with Respect to Marital Status: Additional Results

	Non-Single		Non-Single		Single	
	Married	LT Relationship	Female	Male	Female	Male
	(1)	(2)	(3)	(4)	(5)	(6)
β^{ER}	2.002*	2.345***	2.754***	1.733*	-2.472*	-1.023
	(1.187)	(0.841)	(0.964)	(0.966)	(1.282)	(1.318)
β^{COL}	-2.403**	-0.311	-1.172	-1.366	-0.630	-1.634
	(0.999)	(0.844)	(1.023)	(0.952)	(0.805)	(1.294)
Diff. P-value *[q-value]:*						
ER	0.813 *[0.883]*		0.430 *[0.746]*		0.454 *[0.784]*	
COL	0.109 *[0.640]*		0.509 *[0.919]*		0.890 *[0.746]*	
Pseudo R^2	0.093	0.052	0.079	0.052	0.060	0.027
Observations	259	439	360	338	200	182

Notes: Heteroskedasticity-robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Each column corresponds to a separate Probit regression. Coefficients for Probit regressions of expected rank submission (at baseline) on earnings rank and cost of living (measured by posterior beliefs in baseline survey), and controls (e.g. relative wage, etc.) as defined in section 3. P-values corresponds to the test of the null hypothesis that the coefficients are equal between the two sub-groups. Multiple-testing q-values based on Benjamini and Yekutieli (2001) presented in brackets.

Table D.4: Preference Heterogeneity with Respect to Other Individual Characteristics

	Hypothetical decrease cost of living		Hypothetical increase earnings rank		By Materialism		By Competitiveness		By Life Dimension	
	Better off	Same/Worse off	Better off	Same/Worse off	High	Low	High	Low	High	Low
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
β^{ER}	1.713**	-0.683	1.812*	0.893	1.698**	0.828	1.229*	0.838	1.656*	0.667
	(0.669)	(1.186)	(1.023)	(0.734)	(0.708)	(0.952)	(0.664)	(1.197)	(0.908)	(0.779)
β^{COL}	-1.189**	-1.483	-1.721**	-1.117	-0.638	-2.283***	-1.659***	0.076	-1.028	-1.945**
	(0.585)	(1.159)	(0.799)	(0.710)	(0.757)	(0.751)	(0.608)	(0.958)	(0.748)	(0.869)
Diff. P-value *[q-value]:*										
ER	0.078 *[0.467]*		0.465 *[0.751]*		0.463 *[0.751]*		0.775 *[0.917]*		0.408 *[0.714]*	
COL	0.820 *[0.936]*		0.572 *[0.811]*		0.123 *[0.536]*		0.126 *[0.536]*		0.424 *[0.729]*	
Pseudo R^2	0.043	0.070	0.132	0.033	0.036	0.066	0.043	0.059	0.046	0.063
Observations	782	194	299	677	516	460	750	226	508	468

Notes: Heteroskedasticity-robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Each panel corresponds to a separate Probit regression. Coefficients for Probit regressions of expected rank submission (at baseline) on earnings rank and cost of living (measured by posterior beliefs in baseline survey), and controls (e.g. relative wage, etc.) as defined in section 3. All controls are interacted with indicator variable for heterogeneity variable indicated in panel. In panels c and d, respondents are classified as high/low using the median scores for the competitiveness index (16/30) and materialism index (21/30). Life dimension is based on a principle-component index of rank 5 life dimensions (happiness, health, sense of purpose, spirituality, control over life) that was divided at the median. P-values corresponds to the test of the null hypothesis that the coefficients are equal between the two sub-groups. Multiple-testing q-values based on Benjamini and Yekutieli (2001) presented in brackets.

Table D.5: Robustness to Sample Definition

	Panel A: β^{ER}			Panel B: β^{COL}		
	Non-Single (1)	Single (2)	All (3)	Non-Single (4)	Single (5)	All (6)
Baseline Sample	2.256***	-1.533*	0.990*	-1.087	-1.041	-1.066**
	(0.672)	(0.878)	(0.541)	(0.672)	(0.748)	(0.488)
Pass Attention Check	2.266***	-1.375	1.071**	-0.927	-1.263	-1.079**
	(0.685)	(0.896)	(0.545)	(0.689)	(0.783)	(0.501)
Drop Dual Matches	2.215***	-1.305	1.024*	-1.110	-1.118	-1.133**
	(0.701)	(0.852)	(0.551)	(0.675)	(0.770)	(0.496)

Notes: Heteroskedasticity-robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Raw Probit coefficients. Probit regressions of expected rank order submission on posterior beliefs about cost of living and earnings rank estimated by sample (i.e., coefficients of a same row and sample are from a single regression). All estimates, include the baseline controls listed in section 3. The first row shows estimates for baseline sample (1,080 responses, 698 from non-singles and 382 from singles). The second row restricts the sample to respondents who pass the attention check question in baseline survey (1,041 responses, 678 from non-singles and 363 from singles), while the third row restricts the sample to respondents who are not participating as dual match (1,000 responses, 641 from non-singles and 359 from singles).

Table D.6: Binary Probit vs. Ordered Probit

	Panel A: β^{ER}			Panel B: β^{COL}		
	Non-Single	Single	All	Non-Single	Single	All
	(1)	(2)	(3)	(4)	(5)	(6)
Probit	2.256***	-1.533*	0.990*	-1.087	-1.041	-1.066**
	(0.672)	(0.878)	(0.541)	(0.672)	(0.748)	(0.488)
Ordered Probit	1.362***	-0.356	0.726*	-0.846**	-0.068	-0.560*
	(0.480)	(0.599)	(0.375)	(0.401)	(0.493)	(0.311)

Notes: Heteroskedasticity-robust standard errors in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Raw Probit (or Ordered Probit) coefficients. Probit (or Ordered Probit) regressions of expected rank order submission on posterior beliefs about cost of living and earnings rank estimated by sample (i.e., coefficients of a same row and sample are from a single regression). All estimates, include the baseline controls listed in section 3. All results based on the sample of individuals who completed the baseline survey (1,080 responses, 698 from non-singles and 382 from singles).

Table D.7: IV, First Stage, and Reduced Form Estimates

	Non-Single (1)	Single (2)	All (3)
Panel A: IV-Probit Estimates			
β^{ER}	3.109**	-5.131***	0.910
	(1.360)	(1.942)	(1.165)
β^{COL}	0.387	1.687	0.693
	(1.177)	(1.274)	(0.887)
Panel B: First Stage			
Dep. Var.: $ER^i_{1,2}$			
$\Delta ER^i_{1,2}$	0.853***	0.697***	0.800***
	(0.055)	(0.083)	(0.045)
$\Delta COL^i_{1,2}$	0.017	0.007	0.011
	(0.049)	(0.064)	(0.039)
Dep. Var.: $COL^i_{1,2}$			
$\Delta ER^i_{1,2}$	-0.103***	0.043	-0.057
	(0.036)	(0.091)	(0.038)
$\Delta COL^i_{1,2}$	0.900***	0.984***	0.932***
	(0.064)	(0.070)	(0.048)
Wald test of exog. p-val.	0.308	0.003	0.056
Cragg-Donald F-stat.	169.38	42.37	204.04
Panel C: Reduced Form			
$\Delta ER^i_{1,2}$	2.593**	-3.714**	0.696
	(1.177)	(1.623)	(0.935)
$\Delta COL^i_{1,2}$	0.532	1.814	0.754
	(1.095)	(1.376)	(0.858)
Observations	639	327	966

Notes: Heteroskedasticity-robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Raw Probit (or IV-Probit) coefficients. All regressions include the baseline controls listed in section 3. The independent variables are the posterior beliefs about cost of living and earnings rank, from the baseline specification. Panel A presents raw IV-Probit estimates using model detailed in section 3.2. Panel B shows the first stage for each independent variable. Panel C shows reduced form Probit estimates.

Table D.8: Preference Estimates from Other Studies with Happiness and Hypothetical Data

Reference	Evidence	Country	Parameters	Source	$\frac{a-b}{b}$
Luttmer (2005)	Happiness	U.S.A.	a=0.361, b=0.296	Column (3) of Table 1	0.22
Ferrer-i-Carbonell (2005)	Happiness	Germany	a=0.456, b=0.226	Column (1) of Table 2	1.02
Clark, Senik and Yamada (2016)	Happiness	Japan	a=0.290, b=0.153	Column (1) of Table 3	0.89
Johansson-Stenman, Carlsson and Daruvala (2002)	Hypothetical	Sweden	b/a=0.35	Page 373	1.85
Yamada and Sato (2013)	Hypothetical	Japan	a=0.048, b=0.022	Column (1) of Table 4	1.18

Notes: Authors calculations based on the regression coefficients reported in the papers.

Table D.9: Preferences Inferred from Happiness

	Panel A: β^{ER}			Panel B: β^{COL}		
	Non-Single (1)	Single (2)	All (3)	Non-Single (4)	Single (5)	All (6)
Baseline	1.545**	-0.135	0.936*	-0.661	0.468	-0.311
	(0.634)	(0.954)	(0.520)	(0.621)	(0.761)	(0.479)
Experimental	3.057***	-2.012	1.694*	0.027	1.388	0.488
	(1.098)	(2.024)	(0.975)	(1.049)	(1.236)	(0.794)
Experimental + Long Term	2.919***	-2.361	1.335	0.415	-1.133	-0.212
	(1.078)	(2.183)	(0.983)	(0.949)	(1.211)	(0.760)

Notes: Heteroskedasticity-robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Coefficients for Probit regressions of variable indicating that respondent would live happier life at location 1 (at baseline, or at follow-up for "long term") on earnings rank and cost of living (measured by posterior beliefs in baseline survey), and controls (e.g. relative wage, etc.) as defined in section 3. All results based on the sample of individuals who completed the follow-up survey (978 responses, 647 from non-singles and 311 from singles).

Table D.10: Comparison of Characteristics between Experimental Subjects and Online Sample

	Main Experiment *Med. Students*	Auxiliary Experiment *Online Sample*	Difference
Age	27.091	37.476	-10.385***
	(2.725)	(11.980)	(0.350)
% Male	0.481	0.391	0.090***
	(0.500)	(0.488)	(0.021)
% Married	0.240	0.461	-0.221***
	(0.427)	(0.499)	(0.019)
% Has children	0.089	0.527	-0.438***
	(0.285)	(0.499)	(0.017)
Observations	1,080	1,245	

Notes: Standard deviations and Heteroskedasticity-robust standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Sample statistics for Main Experiment (Medical Student sample) and Auxiliary Experiment (sample of online respondents on Amazon Mechanical Turk).

Table D.11: Experimental Variation, Learning, and Perceptions in Online Sample

			Quality of					Amount of	Percentage of	
Dep. Var.:	$ER_{1,2}^{post}$	$COL_{1,2}^{post}$	Schools	Health	Public Spaces	Environment	Entertainment	Crime	College Graduates	Vote Trump
$\Delta ER_{1,2}$	0.741	-0.095	0.386	-0.066	0.341	0.041	0.077	-0.164	0.574	-0.070
P-value	0.000	0.048	0.000	0.539	0.000	0.689	0.477	0.112	0.000	0.585
Q-value	0.000	0.117	0.001	0.674	0.001	0.783	0.636	0.231	0.000	0.716
$\Delta COL_{1,2}$	0.039	1.017	0.059	-0.136	0.050	0.238	0.092	-0.384	0.001	0.235
P-value	0.513	0.000	0.516	0.139	0.537	0.008	0.349	0.000	0.995	0.033
Q-value	0.674	0.000	0.674	0.273	0.674	0.026	0.504	0.000	0.995	0.091

Notes: Heteroskedasticity-robust standard errors in parenthesis. All regressions include the baseline controls listed in section 3 with the exception of program characteristics. Multiple-testing q-values based on Benjamini and Yekutieli (2001) presented.