

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 November 2020

We are thankful for comments from several colleagues and seminar discussants. 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). AEA RCT Registry #0004203. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

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 They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Nicolas L. Bottan and Ricardo Perez-Truglia. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Choosing Your Pond: Location Choices and Relative Income
Nicolas L. Bottan and Ricardo Perez-Truglia
NBER Working Paper No. 23615
July 2017, Revised November 2020
JEL No. D62,D91,I31,Z13

ABSTRACT

Do individuals care about their relative income? While this is a long-standing hypothesis, revealed-preference evidence remains elusive. We provide a unique test by studying residential choices: individuals often must choose between places with different income distributions, and as a result they “choose” their relative income. We conducted a field experiment with 1,080 senior medical students who participated in the National Resident Matching Program. We estimate their preferences by combining choice data, survey data on perceptions and information-provision experiments. The evidence suggests that individuals care about their relative income and that these preferences differ across single and non-single individuals.

Nicolas L. Bottan
Policy Analysis and Management
Cornell University
MVR Hall 3220
Ithaca, NY 14853
nicolas.bottan@cornell.edu

Ricardo Perez-Truglia
Haas School of Business
University of California, Berkeley
545 Student Services Building #1900
Berkeley, CA 94720-1900
and NBER
ricardotruglia@berkeley.edu

A data appendix is available at <http://www.nber.org/data-appendix/w23615>

1 Introduction

Do individuals care about their relative income? Economists have been pondering this possibility since the seminal work of Adam Smith (Luttmer, 2005). Consistent with the hypothesis that they do indeed care, a large literature shows that subjective well-being increases with relative income (Senik 2004; Luttmer 2005; Ferrer-i Carbonell 2005). The revealed-preference evidence, however, remains elusive. In this paper, we provide unique revealed-preference evidence based on the choice of reference groups.

One infrequent but important choice individuals must make is where to live. By choosing where to live, individuals choose their reference group: that is, the other individuals with whom they will compare their incomes. If individuals choose to move to a poor area, they will be relatively richer than their peers. If they choose to move to a rich area, they will be relatively less rich than their peers. In this paper, we estimate preferences for relative income by studying the location choices of 1,080 senior medical students from the United States.

The identification of preferences for relative income, or any other city amenity, faces a number of data and identification challenges.¹ We developed a novel methodology to address those challenges. Our methodology combines choice data, survey data and an information-provision experiment. While our methodology can be applied to any sample of individuals who are choosing where to live, we implemented it in the specific context of the National Resident Matching Program (NRMP).

The NRMP uses an algorithm to pair graduating medical students with postgraduate hospital residency programs based on rankings submitted by both students and hospitals.

¹For example, one challenge consists of observing not only the city and amenities that the individual chose, but the cities and amenities that the individual could have chosen instead. Some datasets, such as the National Survey of Families and Households (Luttmer 2005) and the United States Postal Service's National Change of Address database (Perez-Truglia 2017), identify individuals moving from one city to another, but do not identify the other cities that the individual could have chosen or incomes they could have earned in those other locations. Another challenge is that we need sources of exogenous variation to identify the amenity preferences. In other words, preferences for a given amenity could be spuriously driven by preferences for another amenity that is not included in the regression.

When submitting their program rankings, the students are choosing the city where they will live for the duration of the residency (typically around five years). Several features of the NRMP make it desirable for this type of revealed-preference analysis (Benjamin et al., 2014). First, the NRMP makes it possible to identify and specify the entire choice set faced by these individuals. Second, because most students are aware of the incentive-compatible matching algorithm used by the NRMP, it is possible to infer preferences directly, without the need for estimating models that rely on additional assumptions. Third, this is a high-stakes choice, to which participants devote ample time and attention. A fourth and convenient aspect of our setting is that the graduating students choose between programs that offer almost identical nominal incomes, but in cities with largely different income distributions and costs of living. As a result, our subjects face substantial tradeoffs between relative income and cost of living. Finally, this context provides a stringent test for estimating preferences for relative income: aspects such as residency quality and prestige are likely the most predominant in medical student's decisions.

We conducted the experiment with 1,080 senior medical students who participated in the 2017 residency match. We conducted a baseline survey roughly about one month before these participants had to submit their rank order preferences. We asked participants about their two most favorite residency programs. Even though students can rank several programs, the ranking of their top two is, in practice, the most important aspect of their decision. We elicited perceptions about their expected position in the city-wide distribution of individual earnings (i.e., their relative income). Additionally, we elicited perceptions about another amenity that is thought to be closely related: their expected cost of living. We then elicited each subject's expected rank submission. Using these data on perceptions and choices, we can estimate how differences in the cities' relative incomes and costs of living affect location choices.

These estimates of location preferences may be subject to omitted variable biases. We use two strategies to address this identification challenge. The first strategy relies on con-

trolling for a host of alternative factors such as the reputation of the residency programs and all the typical city amenities that have been studied in the urban economics literature. The second strategy relies on an information-provision experiment. Immediately after eliciting perceptions about cost of living and earnings rankings, we provided all individuals with statistics about attributes of the cities where their programs are located. 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 who expected to earn \$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 (i.e., their posterior beliefs). This source-randomization experiment creates exogenous variation in posterior beliefs. Our research design exploits that exogenous variation to estimate the causal effects of the perceived attributes on choice by means of an instrumental variables model.²

Our motivation for studying relative income lies in models of social status and consumption aspirations. According to these models, the income of neighbors can impose a negative externality, thus giving individuals an incentive to locate into less affluent areas (Luttmer, 2005). However, there may be other reasons why individuals care about the income of their neighbors. Moreover, some individuals may see the income of neighbors as a positive externality, thus giving them an incentive to move into more affluent areas. For example, certain individuals may expect to have better dating prospects in more affluent areas. Alternatively, individuals may expect to benefit from public goods provided by richer neighbors.

To maximize statistical precision, our baseline estimates exploit all the variation in expectations, experimental and non-experimental. These estimates suggest that, holding cost of living constant, the average individual would prefer to live in a city where her relative

²Our design differs from the traditional approach used in information experiments that consists of using a single source of feedback, but randomly showing feedback to half of the subjects. Instead, we provide information to *all* subjects and randomly nudge some subjects to perceive that they will be richer through the source-randomization.

income would be higher. More precisely, a 1-percentage-point increase in earnings rank, on average, increases the probability that a program will be chosen by 0.186 percentage points (and implying a behavioral elasticity of 0.186).

We use cost of living as a benchmark for relative income. Our baseline estimates suggest that a 1-percentage-point decrease in cost of living, on average, increases the probability that a program will be chosen by 0.201 percentage points (and implying a behavioral elasticity of 0.201). These estimates imply that the average individual would be willing to accept an increase of 0.925 percentage points in the cost of living in exchange for a 1-percentage-point increase in earnings rank. This trade-off suggests that the preferences for relative income are economically significant.

The importance of financial aspects such as relative income and cost of living is consistent with survey data indicating that students chose medical school because of the financial rewards (Daniel and O'Brien 2008). Additionally, these findings are consistent with survey data indicating that, when choosing among programs, applicants care a lot about the geographic location of the programs (Sledge, Leaf, and Sacks 1989). Indeed, when asked to mention the most important factors they consider in their decision-making, 52% of surveyed applicants explicitly mention cost of living (NRMP 2015). Although the preferences for relative income and cost of living are statistically and economically significant, they do not imply that they are the main concern for medical students. Indeed, using perceptions about other program characteristics, we find that doctors care substantially more about prestige and career prospects, which is consistent with the top factors mentioned by medical students in surveys (NRMP 2015).

We show that these average preferences mask meaningful heterogeneity by relationship status. While non-singles prefer being a big fish in a small pond, single individuals prefer being a small fish in a large pond. This difference in preferences is large in magnitude and highly statistically significant (p -value < 0.001). This heterogeneity is consistent with evidence from the happiness literature: Luttmer (2005) finds that the positive effects of

relative income on happiness are driven entirely by non-singles. On the other hand, when it comes to preferences for cost of living, individuals have similar preferences regardless of whether they are single or non-single.

One important caveat with the baseline estimates is that relative income may be picking up the effects of omitted variables. We attempt to address these concerns with two strategies. The first strategy shows that the estimates remain similar after controlling for a host of other characteristics of the residency programs and the cities where they are located. In the second strategy, we focus on the variation in beliefs that was generated by the information-provision experiment. The estimates are less precisely estimated, because we are focusing on a minority of the variation in beliefs, but the main results remain robust.

However, even the experimental estimates are subject to interpretation: individuals may react to the information about relative income not because they care about relative income per se, but because it serves as a signal for other unobservable attributes of a city. While we cannot totally rule it out, we provide two pieces of evidence against this interpretation. We show that individuals are not using the information on earnings rank to update their expectations about cost of living, both in the short term (the baseline survey) and in a longer horizon (the follow-up survey). Second, we conducted an auxiliary experiment with a sample of respondents recruited through Amazon Mechanical Turk. In this complementary experiment we provide feedback about earnings rank and cost of living, but we measure posterior beliefs not only about earnings rank and cost of living but also about other characteristics of the city that the individuals may be learning about based on the earnings rank information (e.g., public goods, crime rates). We show that the results are, again, unchanged when we control for this host of posterior beliefs.

Our favorite explanation for the preference for less affluent ponds – though it is by no means the only explanation – is that individuals see their neighbor’s income as a negative externality. Moving to a more affluent pond, for instance, may result in a loss of social status (Luttmer 2005; Bénabou and Tirole 2006; Ray and Robson 2012), a loss of self-esteem

(Festinger 1954), higher consumption aspirations (Frank 1985b), and poorer outcomes in social interactions (Doob and Gross 1968; Fennis 2008; Nelissen and Meijers 2011). Regarding the heterogeneity by relationship status, our preferred interpretation is that single individuals expect more positive externalities from affluent neighbors such as enjoying public goods and having better dating prospects.

Our paper is related to several strands of literature. First, it relates to literature on the importance of relative income for subjective well-being. Since the seminal contribution of Easterlin (1974), several studies have argued that subjective well-being depends on relative, rather than absolute, income (Clark and Oswald 1996; Senik 2004; Luttmer 2005; Ferrer-i Carbonell 2005; Perez-Truglia 2016).³ Luttmer (2005), for instance, uses data from the United States to show that, holding own income constant, happiness increases with relative income in the area of residence. We contribute to this literature in two ways. Our main contribution is to develop a method to test the relative income hypothesis by using revealed-preference data instead of happiness data. Additionally, even if we take as granted that the income of neighbors imposes a negative externality on happiness, it is unclear whether individuals anticipate these externalities when deciding where to live. Our results provide suggestive evidence that some individuals may anticipate these externalities, at least partially.

This paper is related to a strand of literature using laboratory experiments to study positional externalities. Some of these studies use surveys that let subjects choose between pairs of hypothetical scenarios that encompass tradeoffs between absolute and relative standings. These studies find that individuals are often concerned about their relative standing (Solnick and Hemenway 1998; Johansson-Stenman, Carlsson, and Daruvala 2002; Yamada and Sato 2016; Clark, Senik, and Yamada 2017). In a similar spirit, some laboratory experiments show that relative standing affects behavior in laboratory games such as choices under risk

³These 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, Westergård-Nielsen, and Kristensen 2009). For an extensive review of the literature, see Tideman, Frijters, and Shields 2008.

(Kuziemko, Buell, Reich, and Norton 2014). We contribute to this literature by estimating the tradeoffs between absolute and relative income in a real-world, high-stakes context.

This study is also related to a strand of literature on social status (Frank, 1985a; Heffetz, 2011). For example, Bursztyn, Ferman, Fiorin, Kanz, and Rao (2017) conducted a field experiment to show that individuals use platinum cards to signal their social status. This literature, however, takes reference groups as exogenous and fixed. As a result, there is little direct evidence on how individuals select into reference groups. If individuals truly care about their relative income, when given the choice, they should choose to live in places where their relative income is expected to be higher (Frank, 1985a). Since this is a basic prediction of models of social status, it is notable that this key prediction of models of status has not been tested before. Our contribution is to fill this gap in the literature.

Our paper is additionally related to a large literature measuring preferences for city amenities. Starting with the model pioneered by Rosen (1979) and Roback (1982), economists have been using structural models to estimate the value of amenities (Albouy 2008, 2016). However, there is a growing demand for credible sources of identification. A few studies have exploited quasi-experimental sources of identification to estimate the value of specific amenities. Black (1999), for example, exploits a geographic discontinuity in school-district borders in order to estimate preferences for school quality (see also Bayer, Ferreira, and McMillan 2007). However, due to the limited availability of natural experiments, the quasi-experimental evidence is limited to a small number of amenities and contexts. The methodology developed in this paper can be used to estimate preferences for other types of amenities and contexts. This methodology can provide credible identification of preferences without the need for natural experiments.

Finally, our study is also related to a strand of literature finding that individuals substantially misperceive their relative incomes (Cruces, Perez-Truglia, and Tetaz 2013; Karadja, Mollerstrom, and Seim 2017).⁴ This literature shows that correcting these misperceptions

⁴Misperceptions have been reported on other aspects of society too, for example about the degree of income inequality Hauser and Norton (2017).

has significant effects on stated preferences for redistribution. Yet there is no evidence that these misperceptions have a significant effect on behavior. We contribute to this literature by filling that gap.

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 results on location preferences.

2 Survey Design

2.1 Timing of the Surveys

Inspired by Benjamin et al. (2014), we study the decisions of medical students participating in the NRMP.⁵ After graduating from medical school, students have to complete a residency to become a Medicinæ 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.⁶ After all interviews were completed, the students spend almost two months deciding how to rank their favorite programs. 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, and they may even visit the cities again. In the 2017 Match, the submission window for

⁵Benjamin et al. (2014) 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. They 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. Our survey does differ in some important aspects: we collected our baseline survey *before* subjects submit their rank choices to the NRMP and embedded an information-provision experiment into the survey.

⁶In 2015, the median number of applications submitted was 30 and the median number of interviews 16 (NRMP 2015).

rank order lists opened on January 15 and closed on February 22. We conducted a baseline survey weeks before the submission deadline, which we describe first, and a follow-up survey right after the submission window closed.

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 C.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 earnings rank and the cost of living in the cities where these two programs are located.
3. Feedback: Provide subjects with feedback related to their perceptions.
4. Posterior Beliefs: Re-elicite perceptions about the earnings rank and the cost of living.
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. While in theory students could list any of the residency programs in the country, in practice they tend to focus on the programs that they think may be interested in them. For example, if one student has a flyout at a particular program then that is a strong reason to

list it in the ranking.⁷ Medical students are mostly decided on a specialty when submitting their rankings (indeed, only 1.6% of our sample indicate programs in different specialties). 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 happens 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.⁸

2.2.2 Perceptions about Earnings Rank and Cost of Living

One important feature of the residency match process is that salaries are relatively homogeneous across the different programs, even across specialties.⁹ 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 earnings rank (one for each metro area) and two questions

⁷For more details on the supply and demand in this labor market, see for example Roth and Peranson (1999) and Agarwal (2015).

⁸When 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 6% 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. In Appendix A.9 we show that the results are robust if we exclude those individuals from the sample.

⁹Even though there are no large income differences in residency salaries, there can be large differences in post-residency salaries, especially across specialties. However, this is not a concern because medical students are choosing programs within a specialty. Furthermore, the differences in students' perceptions of post-residency salaries among different programs within specialty is small (Benjamin et al., 2014).

about the cost of living (one for each metro area). We provided an introduction for the question about earnings rank: “Now we want to 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.¹⁰

For the cost of living question, we stated: “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 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.¹¹ In practice, individuals may care about their ranking in finer reference groups: e.g., they may care disproportionately about their relative

¹⁰We decided not to incentivize any of the elicitation – for details, see the discussion in Appendix A.6.

¹¹Moreover, this geographic definition of reference 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.

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 earnings rank and cost of living 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 earnings rank in the two cities being considered and another page with statistics about the cost of living in each of the two cities. 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 60.3% of Champaign-Urbana, IL’s population.” 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.” 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.¹²

After individuals finished reviewing the feedback, we re-elicited their perceptions about these attributes for each city, which we denote as the posterior beliefs. Given that our feedback entailed many figures for participants to remember and process, we wanted to

¹²The median time spent on the feedback page was 18.5 seconds.

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.”¹³

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 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. 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.¹⁴

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

¹³The 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.

¹⁴For more details, see Appendix A.1.

initions. For instance, the earnings rank data is subject to sampling variation because the estimates are based on a limited number of survey respondents, and cost of living data is subject to sampling variation because it tracks the prices of a limited number of goods and services. The variation in definitions arise 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, and different cost of living indices give different weights to expenditure categories.

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. Indeed, we find that the reaction of individuals to the information was orthogonal to the name of the information source.¹⁵

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 (results reported in Appendix A.17).

It is plausible to assume that the rank choices provide a great proxy for the individuals’ true preferences. First, the individuals have strong incentives to take the submission seriously, as whatever they submit is quite consequential: almost all NRMP participants receive a match,¹⁶ and backing out from a match entails serious sanctions.¹⁷ Second, the algorithm

¹⁵Results reported in Appendix A.7.

¹⁶For instance, 95% of the 27,048 U.S. graduating medical students received a successful match in 2017.

¹⁷For example, applicants with confirmed violations of NRMP policies are subject to a one year bar from

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. However, even if the mechanism is strategy-proof in theory, that does not imply that the participants are reporting the truth in practice – for example, they may misunderstand the rules of the mechanism. One nice feature of the setting is that the NRMP provides training to the students making it explicit that it is in their best interest to report preferences truthfully. Moreover, there is some evidence indicating that the training works: for example, survey data from Benjamin et al. (2014) indicates that only around 5% of participants attempt to misreport their true preferences with a strategic motive (see also NRMP 2015 and Rees-Jones 2017).¹⁸ In other words, while the rankings submitted to the algorithm may not be exactly their true preference, they are likely a close approximation.

2.3 Follow-Up Survey

Shortly after the NRMP rank submission window closed, we conducted a follow-up survey with the subjects who responded to the baseline survey. Appendix C.2 shows the full questionnaire of the follow-up survey. Most important, at the 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 earnings rank and cost of living, which allows us to measure the persistence of the information learned in the information-provision experiment. Also, we measured additional characteristics of the subjects that we were not able to measure in the baseline survey due to space and time constraints, such as the places where the individuals grew up.

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.

¹⁸There is a growing literature on how strategy-proof mechanisms work in practice – see for example Li (2017) and Rees-Jones and Skowronek (2017).

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 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:

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

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.¹⁹ 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

¹⁹The 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.

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 very similar if we use alternative models such as Binary Logit, Ordered Probit or Linear Probability (results reported in Appendix A.16).

The two key parameters of interest are β^{ER} and β^{COL} . The parameter β^{ER} measures preferences for relative income during the residency. Depending on the mechanism at play, we may expect β^{ER} to be positive or negative. On the one hand, if individuals see their neighbor’s income as a negative externality, as in the models of social status, then we would expect $\beta^{ER} > 0$ (i.e., individuals want to choose less affluent ponds). On the other hand, if individuals see the neighbor’s income as a positive externality, then we would expect $\beta^{ER} < 0$ (i.e., individuals want to choose more affluent ponds). As our benchmark, the parameter β^{COL} measures preferences for purchasing power during the duration of the residency. The hypothesis is that $\beta^{COL} < 0$: i.e., individuals prefer to live in places with lower cost of living.

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 seven years.²⁰ Note that earnings rank and cost of living *after* the end of the residency would be part of the error term.

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 $\Delta ER_{1,2}^i$

²⁰A small minority of subjects may expect to continue living in the same city after the residency, in which case the distribution of earnings and the cost of living may also be relevant for the post-residency period.

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:

$$I(\text{Program}_1 \succ_i \text{Program}_2) = I(\beta^{ER} \cdot ER_{1,2}^{i,posterior} + \beta^{COL} \cdot COL_{1,2}^{i,posterior} + \lambda_1 \cdot ER_{1,2}^{i,alt} + \lambda_2 \cdot COL_{1,2}^{i,alt} + \theta X^i + \varepsilon_i \geq 0)$$

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

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

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, $\Delta ER_{1,2}^i$ and $\Delta COL_{1,2}^i$ 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

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 C.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.²¹

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 C.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.²²

The only reason why we excluded individuals who had previously submitted their ranks 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 to 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 list 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 partic-

²¹For details, see Appendix A.3.

²²There are a number of alternative matches for some specialties that have different deadlines than the Main Residency Match.

icipating 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.²³

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.²⁴

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 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 in-

²³The 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.

²⁴Upon 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 A.9, the results are virtually the same if we drop this 3.6% of the sample.

dividuals who did not respond to the follow-up.²⁵ 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 (for more details, see Appendix A.3).

Appendix A.2 provides descriptive statistics about the subject pool and shows that. And consistent with successful random assignment, the characteristics are balanced across treatment groups. 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.

The programs that the students are choosing over provide a broad geographical coverage of the U.S. territory (for more details, see Appendix A.2). This broad coverage results in significant trade-offs between higher relative income and lower cost of living, where we observe significant orthogonal variation between them. To show this, Figure 1.a shows the variation between the earnings rank (according to earning \$54,000 in the ACS) and cost of living (according to the RPP). There is substantial orthogonal variation in the two dimensions. For example, the metropolitan areas of Los Angeles, CA and St. Louis, MO have a similar earnings rank, but are dramatically different in cost of living (difference of around 28 percentage points). On the other hand, Figure 1.b shows the orthogonal variation we use by plotting the pairwise differences in cost of living (according to the RPP measure) vs. the differences in earnings rank (according to the ACS measure).²⁶ The substantial dispersion along the y-axis suggests that there are large differences in cost of living across the pairs of cities that the individuals must choose from. The substantial dispersion along the x-axis

²⁵Results presented in Appendix A.3.

²⁶Using the alternative data sources yield similar results.

suggests that there are large differences in earnings rank across the pairs of cities.²⁷ Most important, the $R^2 = 0.22$ indicates that, even though the two are correlated,²⁸ substantial orthogonal variation exists between relative income and cost of living.²⁹

5 Results: Prior Beliefs and Learning

5.1 Distribution of Prior Beliefs

To the best of our knowledge, ours is the first paper to measure perceptions about earnings ranks and cost of living 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 statistics from the baseline sources.

A growing literature documents that individuals have substantial misperceptions about their position in the national income distribution (Cruces et al. 2013; Karadja et al. 2017). Consistent with those findings, our data suggests that respondents have substantial misperceptions about their prospective earnings ranks. Figure 1.c plots prior beliefs about earnings rank against our favorite estimates, from the ACS. On average, individuals underestimate earnings ranks by almost 16 percentage points. This finding is consistent with previous studies asking respondents to guess their position in the national income distribution: Karadja et al. (2017) finds that Swedes underestimate their ranking by an average of 18 percentage points, which is close to the bias of 16 percentage points present in our own data. Moreover, the correlation between prior beliefs and our ACS estimates is positive and statistically significant (p-value<0.001), but the R^2 is low (0.029). Expecting individuals to be accurate in predicting the levels of earnings ranks in different cities may be too demanding. Ultimately,

²⁷Furthermore, the vast majority of these differences in earnings rank and cost of living are orthogonal to differences in nominal income – see Appendix A.5 for details.

²⁸The slope of -0.664 suggests that, on average, relatively more expensive cities tend to have a higher distribution of nominal earnings.

²⁹We provide a number of additional checks in the Appendix . Appendix A.10 shows that there is plenty of orthogonal variation between earnings rank and cost of living even when comparing residency programs of similar reputation. Appendix A.9 shows that the results are robust if we exclude individuals from the 3 largest metro areas (New York, Los Angeles and Chicago) or if we exclude individuals from the 25 smallest metro areas (i.e., those with approximately less than 160,000).

the relevant margin for individuals choosing between two cities is the pairwise difference between the earnings ranks between those two places. Thus, we repeat this exercise using pairwise differences instead of levels of the earnings rank. Indeed, the misperceptions are less substantial when we look at these pairwise differences: the corresponding R^2 of the relationship between perceptions and the ACS estimates is 0.15.³⁰

In comparison, individuals are more accurate identifying the cost of living in the cities they are considering. Figure 1.d 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 prior beliefs 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 highly correlated, with an R^2 of 0.550.

5.2 Learning from Feedback

We next examine whether respondents learned from the information we provided. If respondents learn from the information provided, we would expect a positive relation between their perception gaps (i.e., the signal received minus the prior belief) and their revisions (i.e., the posterior belief minus the prior belief); that is, respondents who originally overestimated would revise their beliefs downwards, while those who underestimated would revise in the opposite direction. Moreover, the slope of this relationship can be used to quantify the degree of learning from the feedback: a parameter that takes the value from 0 to 1, indicating the weight assigned to the signal relative to the weight assigned to the prior belief.³¹

Respondents strongly updated their beliefs after being provided with feedback. Figure 2 presents the reduced-form effects of information for earnings rank and cost of living, respectively. Figures 2.a and 2.d present the short-term effect, that is, the revision made by

³⁰Detailed results reported in Appendix A.6.

³¹See Appendix A.7 for more details.

respondents directly after being given the information. The short-term learning rates, given by the slopes reported in these figures, are 0.873 (s.e. 0.011) for the earnings rank and 0.879 (s.e. 0.010) for the cost of living. These two learning rates are statistically significant (both p -values < 0.001), 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 2.b and 2.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 (0.060 for earnings ranking and -0.034 for cost of living) and precisely estimated. Furthermore, in Appendix A.19 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 they would be “richer than 60%” of that city’s population and then later asked about the cost of living in the same city, he or she may report an earnings rank that is closer to 60% 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 2.c and 2.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 earnings rank it is 0.626 (s.e. 0.020), while for cost of living is 0.752 (s.e. 0.016). The persistent effect of our feedback suggests that individuals indeed care about these city features. 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: Location Preferences

6.1 Average Preferences

We start with the baseline specification, which 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 variations of this specification later.

The Probit coefficients are presented in Table 1. Column (1) presents the results for the full sample, while columns (2) through (7) present results by demographic subgroups. 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. To better understand the magnitude of these Probit coefficients, we can transform them into the corresponding marginal effects, where 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).³²

We use preferences over cost of living as a benchmark for preferences for relative income. The estimated β^{COL} from column (1) is negative and statistically significant (p-value=0.027), suggests that the average individual prefers programs with lower costs of living. The corresponding marginal effect indicates that 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). The elasticity for cost of living (-0.201) is similar in absolute value to the elasticity for earnings rank (0.186) – indeed, their difference is statistically insignificant. This finding suggests that individuals care about relative income nearly about as much as they care about cost of living: the average individual would be willing to accept an increase of 0.925 percentage points in the cost of living in exchange for a 1 percentage point increase in earnings rank.

The fact that medical students care about relative income and cost of living 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). Our findings are also consistent with prior survey evidence indicating that NRMP applicants care about the geographic location of the programs. For example, Sledge et al. (1989) surveyed a sample of graduating medical students to elicit the most important features in determining the program they ranked first in their NRMP submission. They allowed subjects to choose from 15 features that included characteristics of the program (e.g., program curriculum) and other characteristics (e.g., geographic location). They found that “geographic location” was rated as the most important feature (selected 63.8% of the time) as well as the most desirable feature (selected 37.6% of the time). Using more recent data, NRMP (2015) also find that geographic location is the most popular feature; additionally, they find that, when asked to mention the most important factors taken into account in their decision-making, 52% of surveyed medical

³²These marginal effects are reported in Appendix A.8.

students explicitly mention the cost of living.

Even though β^{ER} and β^{COL} are statistically and economically significant, they do not imply that relative income and cost of living during the residency are the main features that medical students pay attention to. Indeed, if doctors are motivated by financial rewards, we would expect them to seek a higher post-residency income through more prestigious programs. Indeed, in complementary analysis, we find that, when deciding their NRMP submissions, doctors care more about the prestige of the program and the career prospects than about the relative income and cost of living during their residency. For example, a standard deviation increase in sense of purpose of a program has an effect that is 3.9 times larger than a standard deviation increase in earnings rank (results presented in Appendix A.11). This finding is also consistent with survey data indicating that the program’s prestige is of higher importance than other attributes such as cost of living (NRMP 2015).

Under the assumption that our revealed-preference estimates reflect the same type of neighbor externalities from Luttmer (2005), we can compare the magnitude of relative concerns in this paper to the magnitude reported in Luttmer (2005). This comparison, presented in Appendix A.13, suggests that our estimates imply a somewhat smaller role for relative concerns.³³

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 non-single individuals. Also, evidence from the urban economics literature indicates that single and non-single individuals have different location preferences (e.g., Couture and Handbury 2016; Gautier et al. 2010). With the goal of exploring this form of heterogeneity, we elicited

³³For a more direct comparison between happiness and choice data, our survey collected a question on happiness: “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 the rank choices. 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 A.14.

the relationship status using the same categories as in Luttmer (2005).

Columns (2) and (3) of Table 1 present the results on this form of heterogeneity, by estimating the model separately for singles and non-singles. Column (2) of Table 1 shows the estimates for non-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).³⁴ 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 the small minority of subjects participating in dual matches.³⁵

The coefficients from columns (2) and (3) suggest a large heterogeneity in β^{ER} by relationship status. For non-single individuals, the estimated β^{ER} (2.236) is positive and highly statistically significant (p-value=0.001). For the sample of single individuals, β^{ER} (-1.538) is negative and statistically significant at the 10% level (p-value=0.080). 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 also highly statistically significant (p-value=0.001). And 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 1, our benchmark, 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. In other words, single individuals tend to see more affluent neighbors as a positive rather than a negative externality. One potential explanation is that they expect to benefit from some public goods that are provided by the more affluent neighbors. Another potential explanation lies in the dating market: since more affluent individuals prefer to date more affluent partners (Fisman et al.

³⁴Appendix A.12 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.

³⁵See Appendix A.9 for more details.

2006, Hitsch et al. 2010), this can naturally create a preference for locating in more affluent ponds.³⁶ Indeed, this result is 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, Gautier et al. (2010) find that after getting married, the dating-market benefits no longer matter for them, and couples move out of the city.

It is important to note that the singles in our subject pool are a special group of individuals, and as a result the dating motive in this sample may be stronger than for the general population of singles. First, these subjects are at their prime dating age, and thus may expect to find a long-term partner during their residency. Second, even though their wages during the residency put them near the middle of the national earnings distribution, their expected *post-residency* earnings will place them at the very top of the national earnings distribution. As a result, these subject would probably struggle to meet a partner that can match their permanent income unless they move into one of the most affluent cities.

Columns (4) through (7) of Table 1 explore other forms of heterogeneity based on characteristics measured in the baseline survey. Columns (4) and (5) explore potential differences in preferences by gender: 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). Columns (6) and (7) explores heterogeneity by future income. Even though all these subjects receive a similar income during the residency, they have different expected incomes after they finish their residencies. Then, it is possible that individuals who selected high-earning specialties may be more concerned about cost of living or relative income. To test this hypothesis, columns (6) and (7) split the sample in specialties with above and below median post-residency average income. The β^{ER} is 1.433 for below-median specialties and 0.777 for above-median special-

³⁶Prior evidence suggests that, relative to single men, single women may have a stronger preference for finding more affluent partners (Bertrand et al. 2015; Bursztyn et al. 2017). 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. Results reported in Appendix A.12.

ties, 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).

To address the possibility of false positives due to multiple hypothesis testing, for each p-value reported in Table 1, we report the corresponding q-value in brackets (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 adjustment for multiple hypothesis testing does not change any of the results: most importantly, 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.

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 relative income, both in terms of magnitude and statistical significance.³⁷ Because of the magnitude of the heterogeneity by relationship status, in the remainder of the paper we report estimates for the entire sample as well as broken down by relationship status.

6.3 Robustness Checks: Additional Controls

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

Table 2 presents the robustness checks. 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

³⁷Results reported in Appendix A.18.

or the metro area. The second row corresponds to the baseline specification from Table 1, which includes the six baseline controls listed in Section 3.1. The results in the first two rows of Table 2 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 2 include different sets of additional controls. These sets of controls were selected based on attributes that could potentially be relevant and at the same time may be correlated to the earnings rank. For instance, if the best rated residency programs were located in less affluent cities that could bias our estimated preferences for relative income.

Each row introduces a different set of control variables: 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 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 subjective program characteristics (the subjective rank in prestige, purpose, and prospect, as reported in the follow-up survey).³⁸ Comparing the results across rows of β^{ER} and β^{COL} of Table 2 suggests that these estimates are robust to the choice of control variables, both in terms of statistical significance and economic significance.³⁹

³⁸The results are similar when using additional controls related to the availability and quality of residency programs in the area – results reported in Appendix A.10.

³⁹Of 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 0.873 with no controls to a maximum of 1.199 with subjective program characteristics. However, all of these differences are statistically insignificant. Also, according to the pseudo- R^2 , including these variables increases the explanatory power of our model to some degree. For the full sample, 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).

6.4 Robustness Checks: Experimental Estimates and Persistence

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 check 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 of the information-provision. To make these estimates directly comparable to the long-term effects, in this section we restrict the sample to the 90.6% of subjects who responded to the follow-up survey.

Panel A of Table 3 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 (and thus more consequential) among singles.

Panel B of Table 3 presents the results for β^{COL} . The results from the baseline specification (first row) are qualitatively different from the results in the experimental specification (second row). All the coefficients (for the entire sample, singles and non-singles) become

positive, are imprecisely estimated, and are statistically insignificant. One potential interpretation could be that individuals anticipate that unobservable city amenities translate into higher cost of living, thus reducing the desire to move to areas with lower cost of living (Roback 1982; Albouy 2008). However, this may simply be a spurious finding: 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.⁴⁰ Indeed, consistent with the interpretation of spurious finding, we find a negative and significant β^{COL} when we replicate the experiment with the secondary sample (results reported in section 6.5 below).

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 3. 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.⁴¹

The second robustness test is intended to address potential concerns about spurious effects of the information-provision experiment. By asking individuals questions about the

⁴⁰Also, 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. For instance, it is plausible that the information-provision experiment disproportionately 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.

⁴¹Reduced-form and first-stage estimates are presented in Appendix A.18.

earnings rank and cost of living, the baseline survey makes those aspects more salient, which may make individuals overweight them in their expected choice. However, it must be noted that this salience effect may not necessarily exaggerate the importance of relative income, because it would be expected to inflate both β^{ER} and β^{COL} . Another potential concern is that of experimenter-demand effects: by providing information about earnings rank and cost of living, 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 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.⁴²

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 purely 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.

The third row of Table 3 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 effects of the information. 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 at the 10% level (p-value=0.093) for non-singles, and negative (-5.285) and statistically significant at the 1% level (p-value=0.008) for singles.

⁴²Also, 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.

6.5 Robustness Checks: Replication

In Appendix B we provide an additional robustness check, based on an auxiliary experiment conducted with a sample of respondents recruited through Amazon Mechanical Turk. This auxiliary sample has the disadvantage that respondents are not planning to move anytime soon, so we must study their hypothetical location choices instead of their actual location choices. The auxiliary sample does have some advantages, however: it is more diverse than the primary sample of medical students in many dimensions such as age and occupation, and it is possible to run further experiments at any time. The results from this auxiliary experiment replicate the main findings from the original experiment: we find that the average individual prefers a higher relative income and a lower cost of living; we find that the experimental estimates are similar to the non-experimental estimates; and we find large differences in preference for relative income by relationship status. We also compare the marginal rate of substitution between relative income and cost of living between the auxiliary and main samples. The results indicate that, while somewhat smaller in magnitude, the preference for relative income is still statistically and economically significant in the, more diverse, auxiliary sample.

6.6 Robustness Checks: Alternative Mechanisms

One possible interpretation of the coefficient β^{ER} is that individuals use their prospective relative income as a signal for other city attributes. 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. In any case, we present two tests for this mechanism. One version of this mechanism 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. In Appendix A.19 we show evidence against this confounding factor, by showing that the information about relative income did not affect posterior beliefs about the expected cost of living. A second version of this mechanism is that individuals use information about relative income to learn about other city characteristics, such as school quality and crime rates. In Appendix B.6 we use data from the auxiliary experiment, in which we elicited posterior beliefs about these and other attributes of the city, to show evidence against this confounding factor.

7 Conclusions

We conducted a field experiment with a sample of 1,080 medical students who participated in the National Resident Matching Program. We provide unique revealed-preference evidence that individuals care about their prospective relative income. We find that, on average, individuals want to live in places where their relative income will be higher. This evidence is consistent with the predictions of models of relative concerns, according to which affluent neighbors would impose a negative externality (Luttmer 2005). We also show that there is meaningful heterogeneity in these preferences for relative income: non-single individuals want to live in less affluent ponds, whereas single individuals prefer to live in more affluent ponds. The heterogeneity by relationship status suggests that, relative to non-singles, singles see more affluent neighbors as a positive externality. For example, singles may expect to benefit from dating their affluent neighbors, or to benefit from the public goods that they finance.

Regarding the external validity of our results, it is possible that senior medical students care about their relative income more than the general population. For example, given that most doctors make well above the subsistence level, they may have the luxury of caring about their relative income. Doctors may also be more materialistic (Daniel and O'Brien

2008). Indeed, the results from our auxiliary survey, based on a more diverse cross-section of the U.S. population, suggests that the preference for relative income is still present but somewhat smaller in magnitude for the general population. An 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, even though the settings may not be as clear-cut as for the medical residency, there are multiple job markets in which job seekers must choose between job offers in different cities.⁴³ Future research could also investigate the mechanisms underlying individuals' concerns about relative income. There is, for instance, little evidence as to whether the preference for relative income responds to instrumental (e.g., dating prospects) versus non-instrumental motives (e.g., envy).⁴⁴ Last, we hope our methodology could be used in the future to study not only preferences for relative income and cost of living, but to study preferences over various different city amenities.

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.
- Bayer, P., F. Ferreira, and R. McMillan (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy* 115(4), 588–638.
- Bénabou, R. and J. Tirole (2006). Incentives and prosocial behavior. *The American Economic Review* 96(5), 1652–1678.
- Benjamin, D. J., O. Heffetz, M. S. Kimball, and A. Rees-Jones (2014). Can marginal rates of substitution be inferred from happiness data? evidence from residency choices. *The American Economic Review* 104(11), 3498–3528.
- Benjamini, Y. and D. Yekutieli (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* 29(4), 1165–1188.

⁴³Some examples are: employees on the job market considering job offers, prospective undergraduate or graduate students choosing where to study, graduating students choosing among their jobs after graduation, etc.

⁴⁴These additional hypotheses can be explored with the same empirical framework proposed in this paper, but with additional treatment arms designed to test specific mechanisms.

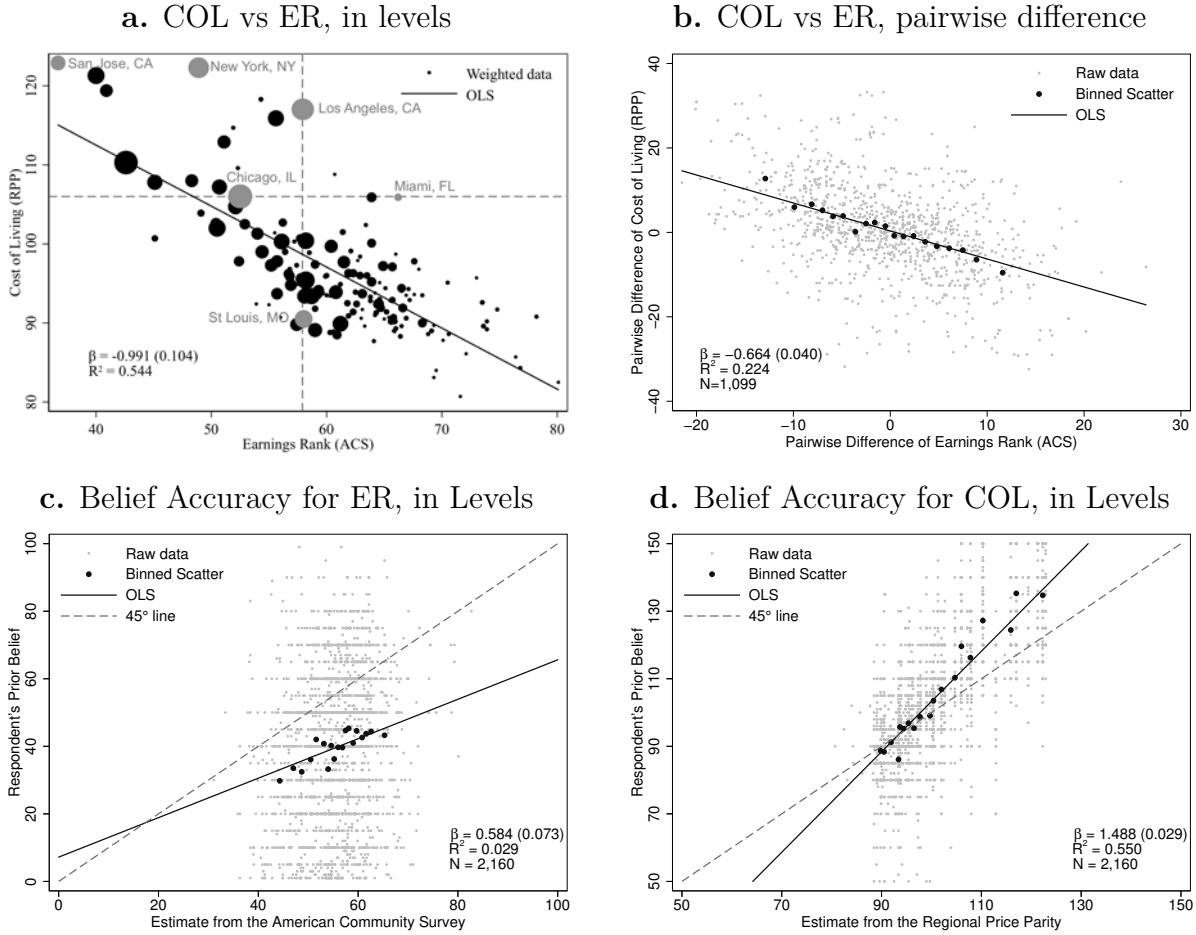
- Bertrand, M., E. Kamenica, and J. Pan (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics* 130(2), 571–614.
- Black, S. E. (1999). Do better schools matter? parental valuation of elementary education. *The Quarterly Journal of Economics* 114(2), 577–599.
- Bursztyn, L., B. Ferman, S. Fiorin, M. Kanz, and G. Rao (2017). Status goods: experimental evidence from platinum credit cards. *NBER Working Paper No. 23414*.
- Bursztyn, L., T. Fujiwara, and A. Pallais (2017). Acting wife: Marriage market incentives and labor market investments. *The American Economic Review* 107(11), 3288–3319.
- Cavallo, A., G. Cruces, and R. Perez-Truglia (2017). Inflation expectations, learning, and supermarket prices: Evidence from survey experiments. *American Economic Journal: Macroeconomics* 9(3), 1–35.
- Clark, A., C. Senik, and K. Yamada (2017). When experienced and decision utility concur: The case of income comparisons. *Journal of Behavioral and Experimental Economics* 70, 1 – 9.
- Clark, A. E. and A. J. Oswald (1996). Satisfaction and comparison income. *Journal of Public Economics* 61(3), 359–381.
- Clark, A. E., N. Westergård-Nielsen, and N. Kristensen (2009). Economic satisfaction and income rank in small neighbourhoods. *Journal of the European Economic Association* 7(2-3), 519–527.
- Couture, V. and J. Handbury (2016). Urban revival in america, 2000 to 2010. *University of Pennsylvania. Mimeo*.
- Cruces, G., R. Perez-Truglia, and M. Tetaz (2013). Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. *Journal of Public Economics* 98, 100–112.
- Daniel, C. and M. O’Brien (2008). Why study medicine? *The Student Doctor Network April 24*.
- Doob, A. N. and A. E. Gross (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.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations* 7(2), 117–140.
- Fisman, R., S. S. Iyengar, E. Kamenica, and I. Simonson (2006). Gender differences in mate selection: Evidence from a speed dating experiment. *The Quarterly Journal of Economics* 121(2), 673–697.
- Frank, R. H. (1985a). *Choosing the right pond: Human behavior and the quest for status*.

Oxford University Press.

- Frank, R. H. (1985b). The demand for unobservable and other nonpositional goods. *The American Economic Review* 75(1), 101–116.
- Gautier, P. A., M. Svarer, and C. N. Teulings (2010). Marriage and the city: Search frictions and sorting of singles. *Journal of Urban Economics* 67(2), 206 – 218.
- Hauser, O. P. and M. I. Norton (2017). (mis)perceptions of inequality. *Current Opinion in Psychology* 18, 21 – 25. Inequality and social class.
- Heffetz, O. (2011). A test of conspicuous consumption: Visibility and income elasticities. *Review of Economics and Statistics* 93(4), 1101–1117.
- Hitsch, G., A. Hortacsu, and D. Ariely (2010). What makes you click: An empirical analysis of online dating. *Quantitative Marketing and Economics* 8(4), 393–427.
- Johansson-Stenman, O., F. Carlsson, and D. Daruvala (2002). Measuring future grandparents’ preferences for equality and relative standing. *The Economic Journal* 112(479), 362–383.
- Kahneman, D. and A. Tversky (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology* 3(3), 430–454.
- Karadja, M., J. Mollerstrom, and D. Seim (2017). Richer (and holier) than thou? the effect of relative income improvements on demand for redistribution. *Review of Economics and Statistics* 99(2), 201–212.
- Kuziemko, I., R. W. Buell, T. Reich, and M. I. Norton (2014). Last-place aversion: Evidence and redistributive implications. *The Quarterly Journal of Economics* 129(1), 105–149.
- Li, S. (2017, November). Obviously strategy-proof mechanisms. *American Economic Review* 107(11), 3257–87.
- 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 M. H. Meijers (2011). Social benefits of luxury brands as costly signals of wealth and status. *Evolution and Human Behavior* 32(5), 343–355.
- NRMP (2015). 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 G. Cruces (2017). Partisan interactions: Evidence from a field experiment in the united states. *Journal of Political Economy* 125(4), 1208–1243.
- Ray, D. and A. Robson (2012). Status, intertemporal choice, and risk-taking. *Econometrica* 80(4), 1505–1531.
- 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 S. Skowronek (2017). Why do we lie in incentive-compatible mechanisms? evidence from the residency match. *Mimeo*.

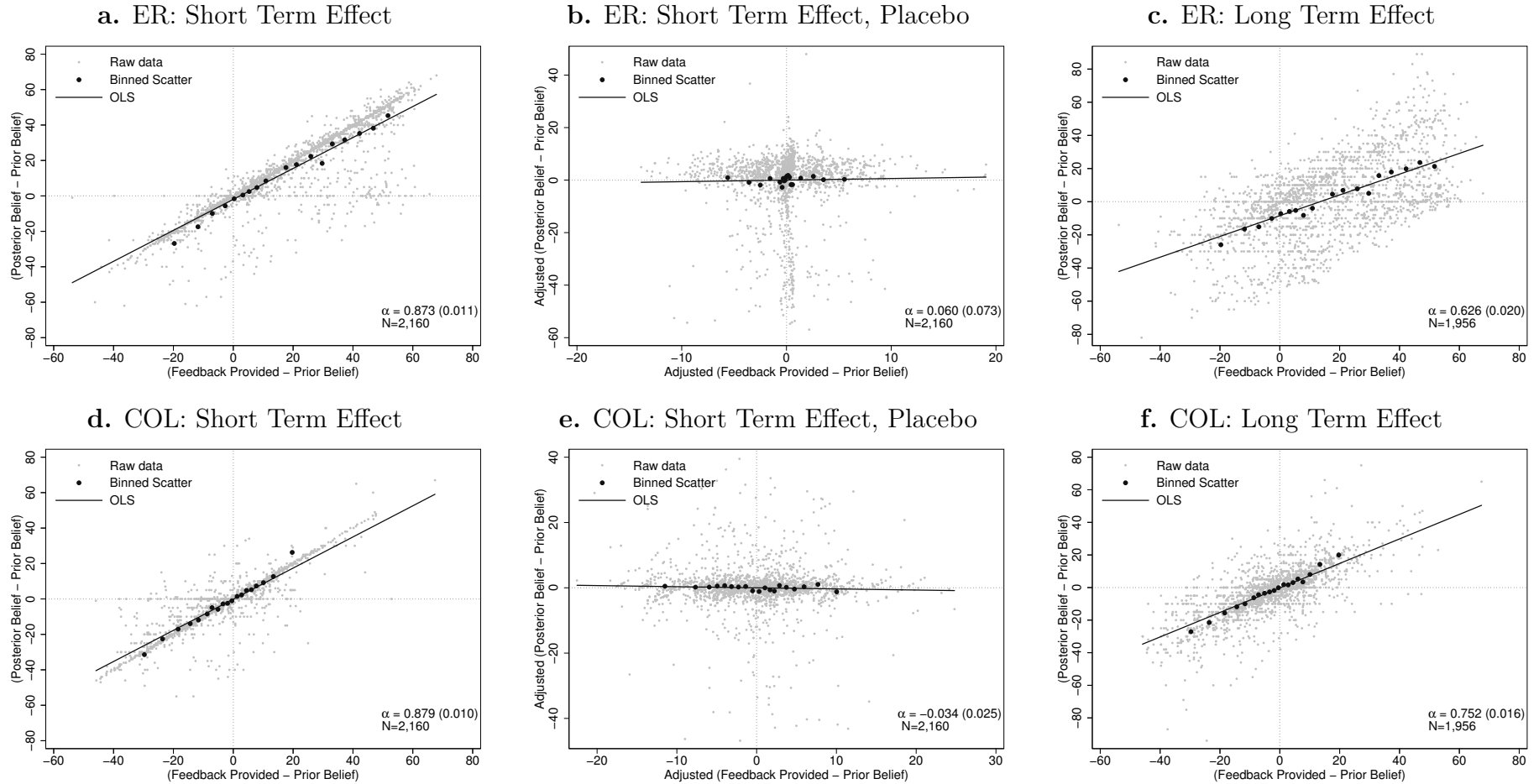
- Roback, J. (1982). Wages, rents, and the quality of life. *Journal of Political Economy* 90(6), 1257–1278.
- Rosen, S. (1979). Wage-based indexes of urban quality of life. *Current Issues in Urban Economics*, 74–104.
- Roth, A. E. and E. Peranson (1999). The redesign of the matching market for american physicians: Some engineering aspects of economic design. *The American Economic Review* 89(4), 748.
- Senik, C. (2004). When information dominates comparison: Learning from russian subjective panel data. *Journal of Public Economics* 88(9), 2099–2123.
- Shigeoka, H. and K. Yamada (2016). Income-comparison attitudes in the us and the uk: Evidence from discrete-choice experiments. *NBER Working Paper No. 21998*.
- Sledge, W. H., P. Leaf, and M. Sacks (1989). Applicants perceptions of psychiatric residency training programs. *Academic Psychiatry* 13(1), 24–30.
- Solnick, S. J. and D. Hemenway (1998). Is more always better? a survey on positional concerns. *Journal of Economic Behavior & Organization* 37(3), 373–383.
- Tideman, S., P. Frijters, and M. A. Shields (2008). Relative income, happiness, and utility: An explanation for the easterlin paradox and other puzzles. *Journal of Economic Literature* 46(1), 95–144.
- Van Gelder, M. M., R. W. Bretveld, and N. Roeleveld (2010). Web-based questionnaires: the future in epidemiology? *American Journal of Epidemiology* 172(11), 1292–1298.
- Yamada, K. and M. Sato (2016). *Another avenue for anatomy of income comparisons: Evidence from hypothetical choice experiments*. Behavioral Economics of Preferences, Choices, and Happiness; Springer.

Figure 1: Variation in Earnings Rank, Cost of Living and Belief Accuracy



Notes: (a) Dots are weighted responses for each metro area chosen by subjects. Variables for x-axis and y-axis correspond to levels of the earnings rank and the cost of living. (b) The gray dots correspond to the raw scatterplot, and the darker dots correspond to the binned-scatterplot based on 20 bins. 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). (c)-(d) 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). Slopes (β , with robust standard errors in parentheses) and R^2 are based on a linear regression.

Figure 2: 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.

Table 1: Location Preferences: Baseline Estimates

	By Relationship Status			By Gender		By Specialty Salary	
	All (1)	Non-Single (2)	Single (3)	Female (4)	Male (5)	\leq \$229,000 (6)	$>$ \$229,000 (7)
β^{ER}	0.995* (0.539)	2.236*** (0.669)	-1.538* (0.880)	1.041 (0.755)	0.896 (0.781)	1.433* (0.732)	0.777 (0.797)
β^{COL}	-1.073** (0.485)	-1.087 (0.663)	-1.058 (0.749)	-0.972 (0.679)	-1.443* (0.753)	-0.690 (0.713)	-1.238* (0.690)
Diff. P-value [<i>q-value</i>]:							
ER		0.001 [<i>0.030</i>]		0.894 [<i>0.974</i>]		0.544 [<i>0.954</i>]	
COL		0.977 [<i>0.977</i>]		0.642 [<i>0.954</i>]		0.580 [<i>0.954</i>]	
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 earnings rank and cost of living, 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 2: Location Preferences: Robustness to Alternative Control Variables

	Panel A: β^{ER}			Panel B: β^{COL}		
	Non-Single (1)	Single (2)	All (3)	Non-Single (4)	Single (5)	All (6)
No Controls	1.961*** (0.663)	-1.480* (0.841)	0.873* (0.531)	-0.812 (0.523)	-1.131* (0.589)	-0.894** (0.382)
Baseline	2.236*** (0.669)	-1.538* (0.880)	0.995* (0.539)	-1.087 (0.663)	-1.058 (0.749)	-1.073** (0.485)
Demographic	2.288*** (0.715)	-0.871 (0.977)	1.176** (0.578)	-1.219* (0.628)	-1.712** (0.713)	-1.342*** (0.468)
Amenities	2.056*** (0.669)	-1.381 (0.853)	0.958* (0.538)	-0.718 (0.630)	-1.265 (0.816)	-0.898* (0.481)
Geography	2.011*** (0.658)	-1.214 (0.869)	1.010* (0.528)	-1.093** (0.547)	-1.393** (0.592)	-1.168*** (0.401)
Economic	1.914*** (0.684)	-1.191 (0.941)	0.946* (0.566)	-0.467 (0.670)	-1.647** (0.812)	-0.868* (0.498)
State Dummies	2.901*** (0.703)	-1.943** (0.907)	1.084* (0.555)	-1.090 (0.671)	-1.219 (0.939)	-0.968* (0.502)
Obj. Program Chars.	2.018*** (0.676)	-1.578* (0.847)	0.904* (0.535)	-0.840 (0.516)	-1.151** (0.587)	-0.910** (0.381)
Subj. Program Chars.	2.222*** (0.730)	-1.320 (1.040)	1.199** (0.605)	-1.210** (0.587)	-1.678*** (0.619)	-1.277*** (0.425)
All Controls	2.385*** (0.887)	-0.685 (1.797)	1.147* (0.684)	-0.550 (1.008)	-4.165*** (1.598)	

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 earnings rank and cost of living, 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 two rows, which are restricted to the follow-up sample (978 responses, 595 from non-singles and 311 from singles).

Table 3: Location Preferences: 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*** (0.702)	-1.656* (0.991)	1.141** (0.577)	-1.234* (0.743)	-1.379* (0.772)	-1.262** (0.531)
Experimental	2.977** (1.331)	-4.964** (1.974)	0.867 (1.151)	0.353 (1.160)	1.663 (1.286)	0.662 (0.881)
Experimental, Long Term	1.993* (1.188)	-5.285*** (1.984)	-0.029 (1.071)	1.662* (1.005)	0.251 (1.359)	1.012 (0.821)
Experimental, Falsification	-0.007 (0.998)	0.040 (1.732)	0.004 (0.837)	0.037 (0.855)	0.021 (1.123)	0.031 (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 earnings rank and cost of living, 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).