Informed Choices: Gender Gaps in Career Advice*

Yana Gallen and Melanie Wasserman

December 2020

Abstract

This paper estimates gender differences in access to informal information regarding the labor market. We conduct a large-scale field experiment in which real college students seek information from working professionals about various career paths, and we randomize whether a professional receives a message from a male or a female student. We focus the experimental design and analysis on two career attributes that prior research has shown to differentially affect the labor market choices of women: the extent to which a career accommodates work/life balance and has a competitive culture. When students ask broadly for information about a career, we find that female students receive substantially more information on work/life balance relative to male students. This gender difference persists when students disclose that they are concerned about work/life balance. In contrast, professionals mention workplace culture to male and female students at similar rates. After the study, female students are more dissuaded from their preferred career path than are male students, and this difference is in part explained by the greater emphasis on work/life balance to female students. Since the experimental design incorporates real students, we are able to elicit students' preferences for professionals and assess whether average gender differences in information provision are attenuated by students' choices of whom to contact. Our findings suggest that gender differences in information provision persist after accounting for student selection of professionals.

Keywords: career information; gender; discrimination; correspondence study

JEL Codes: C93, J16, J24, J71

*Gallen: University of Chicago, Harris School of Public Policy. E-mail: yana@uchicago.edu. Wasserman: UCLA Anderson School of Management. E-mail: melanie.wasserman@anderson.ucla.edu. Funding is gratefully acknowledged from the Upjohn Institute Early Career Research Grant, UCLA Faculty Senate Grant, and the Initiative for the Study of Gender in the Economy at the Becker Friedman Institute at the University of Chicago. We thank Jack Landry, Mikaela Hassenzahl, Amber Liu, Jessica Lyu, Jacob Kohlhepp, and Matias Giaccobasso for excellent research assistance. We have benefited from comments from Dan Black, Paola Giuliano, Romain Wacziarg, Brendan Price, Manasi Deshpande, Trevor Gallen, Jeff Grogger, Juanna Joensen, Matt Notowidigdo, and Nico Voigtlaender, and seminar participants at NBER Summer Institute 2020 Gender in the Economy workshop, University of Maryland Population Research Center, Chicago Federal Reserve Bank, University of Michigan, NBER Labor Studies Fall 2020 Meeting, UC Santa Barbara Applied Micro Lunch, University of Chicago Experimental Brownbag, Advances with Field Experiments, and the University of Chicago Demography Workshop. This study was registered in the AEA Social Sciences Registry under AEARCTR-0005464.

1 Introduction

Information is essential to economic decision-making. A growing literature has demonstrated that providing individuals with information on the returns to education, the earnings associated with college majors, or the attributes of jobs causes individuals to update their beliefs about these key economic parameters and alter their choices (Jensen, 2010; Nguyen, 2008; Wiswall and Zafar, 2013). Information is typically not provided in a research setting, however. Rather, it is often obtained through informal and private means: a conversation with a professor in her office, an interaction with a prospective employer over email, or a text message with a friend (Montgomery, 1991; Duflo and Saez, 2003; Acemoglu et al., 2014). These informal communications may not be readily accessible to all individuals and may be influenced by the characteristics of participants. In this paper, we study gender differences in access to informal information. To the extent there are gender disparities in access to informal information, these differences could help explain broader phenomena such as occupational segregation by gender, gender gaps in investment choices, and gender differences in health behaviors (Bertrand, 2011; Blau and Kahn, 2017; Dwyer et al., 2002; Cawley and Ruhm, 2011).

This paper focuses on informal exchanges between college students and professionals for the purpose of gathering information regarding their career decisions. Specifically, we investigate whether student gender causally affects the information that students receive regarding various career paths. To isolate whether student gender alters the access to and the content of information provided on career paths, we implement a large-scale field experiment that generates informal interactions between students and professionals. We recruit undergraduate students interested in learning about various careers to send messages on an online professional platform and we randomize whether a professional receives a message from a male or a female student. The messages sent by students are pre-formulated questions seeking information about the professional's career path. The use of the professional networking platform as well as the text of the messages are based on a university career center's guidance for informational interviews. We utilize four question templates, each intended to test a specific hypothesis regarding gender differences in information acquisition.

The first question asks broadly about the pros and cons of a given career path, which allows us to test whether professionals emphasize different career attributes to male and female students. The two career attributes we focus on are work/life balance and competitive culture, both of which differentially affect the labor market choices of women (Goldin, 2014; Wiswall and Zafar, 2018; Niederle and Vesterlund, 2011; Flory et al., 2015). One reason professionals may differentially emphasize these career attributes to female students is that the professionals statistically discriminate, that is, believe female students care more about these issues. Using the next two questions, we test whether gender differences in information provision persist when we control for professionals' perceptions of what students care about. These questions ask specifically

for information about work/life balance or competitive culture. A remaining explanation for professionals' differential willingness to discuss these career attributes with female students is that professionals believe female students will experience more challenges with these issues during their careers. The fourth question is intended to control for the potentially different future experiences of male and female students, by testing whether there are gender differences in responses to a fact-based question on the billable hours requirements in one of the career paths.

Our main finding is that the information professionals provide depends on the gender of the student who is asking. While male and female students have similar response rates to the broad question that asks about the pros/cons of the professional's field, the text of the responses reveals substantial gender disparities. Professionals are more than two times as likely to provide information on work/life balance issues to female students relative to male students. The vast majority of these mentions of work/life balance are negative and increase students' concern about this issue. In addition, when students ask specifically about work/life balance, female students receive 26 percent more responses than do male students. This means that the differential emphasis on work/life balance to female students in responses to the broad question is not entirely driven by perceptions that female students care more about this issue. Interestingly, there is no differential emphasis on workplace culture to female students, through either mentions in responses to the broad question or response rates to the specific competitive culture question. To our knowledge, this is the first study to causally isolate how individuals' gender shapes the information they receive about career attributes.

To understand the ramifications of differential information provision in informal interactions, consider the analogous *formal* information provision scenario. Suppose students visit a university career planning office and express interest in a particular career. If gender-specific career guidance pamphlets are available, some female students may choose the male pamphlet, and some male students may choose the female pamphlet.[1] Some students may take both pamphlets. Importantly, the information is available to all students, regardless of gender. Differential information provision – akin to what we find in informal interactions – would imply that female students receive a pamphlet emphasizing certain career attributes, while male students receive a pamphlet emphasizing others, and no student is aware that there are two pamphlets available or that they were allocated based on gender. The lack of awareness regarding information deficits and on what basis information is provided may give rise to long-term disparities in knowledge and outcomes.

We conduct a follow-up survey to investigate whether gender gaps in information received translate into gender gaps in career outcomes. Students are asked whether, relative to the start of the study, they are

[1]Gender-specific articles on careers are commonly available in newspapers, periodicals, trade newsletters, etc. In addition, books such as *Lean In* by Sheryl Sandberg, provide career advice tailored to women.

more or less likely to enter their preferred career path. Results of the survey indicate that female students become less likely to enter their preferred career path relative to male students. Moreover, nearly half of the gender gap in career plans can be explained by the fact that work/life balance is emphasized more to female students than to male students. The deterring effect of work/life balance mentions makes sense given the generally negative content of this information. We view this as suggestive evidence that gender differences in access to information may lead to occupational segregation by gender.

We test for the possibility that different types of professionals respond to male and female students, with varying inclinations to bring up work/life balance. This could occur, for instance, if female professionals are more likely to respond to female students and are also more likely to mention work/life balance issues relative to male professionals. We find that differences in the composition of professionals who respond to male and female students do not explain *any* of the greater emphasis on work/life balance to female students. We also find suggestive evidence that the additional information on work/life balance provided to female students crowds out the provision of other – potentially useful – career information.

Our experimental design resembles a traditional correspondence study in that the researchers maintain control over the format and content of the message sent, and student gender is orthogonal to professionals' characteristics. In contrast to a traditional correspondence study that uses fictitious individuals, our paper incorporates real students interested in career information, who seek advice from real professionals. While incorporating real students cedes precise control over student attributes, we take several steps to ensure "all else is equal." First, we recruit students from similar majors and similar years in college, who are interested in career information in the four broad career paths. Second, we strictly limit other personal, academic, and professional information on the student's profile on the platform to ensure that the students are perceived as otherwise similar, aside from their gender. Third, in the regression specifications we control for all directly observable information on students' profiles. Fourth, we test the robustness of the results to the inclusion of student characteristics that could be inferred from the profile or observed elsewhere online. Finally, we find similar results when we limit the sample to students with no online presence aside from their profile.

The field experiment permits estimation of *average* gender differences in information received. Any average difference may be amplified or attenuated depending on students' preferences for whom to contact. While every professional contacted is a graduate of a top university working in one of the fields of interest to students, we find that students do not prefer all professionals equally. Before sending messages, students are asked to rank the professionals they are most interested in asking the types of questions in the study. We use these rankings to test whether preferred professionals give different answers than the average professional in the study, and whether female students select professionals who are less (or more) gender biased in their responses. After accounting for student selection into informal interactions, we find that responses to female

students are still more likely to mention work/life balance issues. To our knowledge, this is the first paper to address the limitation that correspondence studies only inform us about average discrimination.

This paper contributes to a number of literatures. First, this paper advances the robust literature investigating the role of informational frictions in shaping individuals' perceptions of key economic variables used in decision-making. Narrowing the discussion to papers that pertain to educational and occupational decisions, this literature generally tests the effects of information *provision* on beliefs regarding the returns to education (Jensen, 2010; Nguyen, 2008; Dinkelman and Claudia Martínez, 2014; Hoxby and Turner, 2015), academic majors (Zafar, 2011; Wiswall and Zafar, 2013, 2015), and occupations and jobs (Coffman et al., 2017; Wiswall and Zafar, 2018). These papers establish that individuals, on average, are misinformed regarding fundamental economic parameters that guide decision-making, such as the earnings associated with various college majors. Furthermore, individuals update their beliefs, stated choices, and actual choices when presented with accurate information on these parameters. Our paper provides a novel advance by investigating access to information. Specifically, we test whether there are gender differences in access to informal information about careers that can potentially contribute to disparities in expectations, preferences, and decisions regarding career paths.

Second, we contribute to the nascent literature that tests whether there is differential treatment of men and women in information-seeking settings. The most closely related papers are correspondence studies by Milkman et al. (2015) and Kalla et al. (2018). Milkman et al. (2015) uses fictitious prospective Ph.D. students to send emails to faculty members asking about research opportunities, and finds that women and minorities are less likely to receive a reply than white men. The messages used in the study could be interpreted as seeking employment, information about employment, or both. Our study omits the job-seeking motive of messages, by emphasizing that the student is not currently looking for job opportunities. Furthermore, our study analyzes the content of responses as well as response rates. Kalla et al. (2018) implements a large-scale experiment that uses fictitious students to send emails to local politicians seeking advice for a class project on how to become a politician. The study finds men and women are equally likely to receive a response. Our paper adds to this literature by focusing on gender differences in access to basic information about various career attributes. In addition, the messages are sent by real students, which emulates a realistic interaction that would occur on the professional platform, and allows us to explore the role of selection.

Finally, this paper provides two advances to the literature that relies on correspondence studies to estimate discrimination. In a traditional resume study designed to estimate the effect of job applicant characteristics on callback rates, fictitious resumes with randomized applicant characteristics are sent to employers. One issue that has been raised regarding these studies is that – due to the fictional nature of the job applicants – employers are being deceived and their time is being wasted (Pager, 2007; Bertrand and Duflo, 2017). An

employer will never be able to hire one of these applicants. Our paper resolves this issue by incorporating real students interested in career information, seeking advice from real professionals. The cost of incorporating real students is ceding control over some student characteristics. While we believe the online setting mitigates the concern that other student characteristics confound the effect of student gender, we test the sensitivity of our results to gender differences in observable and unobservable characteristics of students.

Because audit and correspondence studies rely on randomization to estimate discrimination, the estimates are only informative of discrimination on average (Heckman, 1998; Bertrand and Duflo, 2017). A further innovation of the present paper is to incorporate the fact that, in reality, students are unlikely to contact professionals at random. We incorporate information on student preferences for the professionals they want to interact with to quantify the role of selection into informal interactions in attenuating or amplifying average differences in information provision.[2] The role of selection is potentially important in other settings, such as the labor market, retail market, dating market, and credit market. Future correspondence studies might consider incorporating the search behavior of real individuals to adjust their estimates of average discrimination.

The paper proceeds as follows. In Section 2 we detail the experimental design. Section 3 describes the data collected and presents summary statistics. Section 4 reports the main results on gender differences in information received. Section 5 investigates the implications of gender differences in information provision for students' career choice. Section 6 assesses various mechanisms for the gender differences. In Section 7 we combine the results of the field experiment with student preferences to determine whether average gender differences in information received are amplified or attenuated by student selection of whom to contact for information.

2 Experimental design

To isolate the causal effect of student gender on the information students receive regarding careers, we implement a large-scale field experiment, in which college students solicit information from professionals on an online professional networking platform.

Process

From February 2020 to June 2020, we recruited 100 undergraduate students at a large research university to send messages to professionals. We advertised the study using email lists for the undergraduate economics,

[2]To our knowledge, this paper is the first to incorporate agent selection into a correspondence-style study. A review of 80 audit and correspondence studies yields no other paper that accounts for agent selection.

public policy and math majors, extracurricular clubs related to economics, and undergraduate economics courses. The advertisement was targeted to students interested in career advice. Students interested in participating were asked to fill out a background survey, in which we asked for basic demographic information as well as whether the student was interested in receiving information on four career paths that undergraduate economics majors commonly choose post-graduation: finance, management consulting, data science, and law. We selected students who expressed interest in receiving information on the career attributes of these fields.

In an in-person or virtual meeting, each student participant was guided through the process of creating a profile on an online professional networking site.[3] We asked that each student restrict their profile to minimal information, including their first name and last initial, student status, university affiliation, start year and anticipated year of graduation, college major, and the number of network connections they have on this platform. Students who already had a profile were asked to temporarily remove other information from their profile for the three-week duration of the study. We provided students with the same photo of an iconic university building to use as a profile picture. We confirmed that students created a profile with the requisite restrictions through profile screenshots and independent verification on the platform.

The pool of professionals consists of approximately 10,000 individuals on the online professional platform with work experience in the fields of finance, management consulting, law, or data science. The professionals were found through a search of the professional networking site for individuals who work in the students' metropolitan area, who have work experience in at least one of the four fields, and who have a degree from a U.S. News and World Report top 40 ranked university. Professionals' profiles were checked to ensure they have work experience in one of the four fields.

Each student was given a list of 100 professionals to contact: 13 data scientists, 28 finance professionals, 33 lawyers, and 26 management consultants. These proportions reflect the composition of professionals that came up in a search of the online platform. Each professional received at most one message. We provided the text of the initial message that students sent to professionals. Each professional-student communication used one of four message types, which were designed to emulate a conventional request for career information during an informational interview.[4] To test whether professionals emphasize different career attributes to male and female students, the broad message asked about the pros and cons of the professional's field. To test whether men and women receive different advice conditional on raising a particular concern, we sent two message types that ask whether specific career attributes are concerns – work/life balance and competitive culture. We selected these concerns based on documented gender differences in preferences for competitive environments and temporal flexibility (Goldin, 2014; Wiswall and Zafar, 2018; Niederle and Vesterlund, 2011;

[3]Instructions for this process are available upon request.

[4]These messages were based on suggested wording from a university career center guide on informational interviews. See pages 10 and 11 of `https://career.ucla.edu/Portals/123/documents/career%20guide/UCLA_CareerGuide_2019-2021.pdf`.

Flory et al., 2015). Last, to test whether there are differences in the factual content of advice given to men and women, we asked about the billable hours requirement for a lawyer at a large law firm. This message type was only sent to law professionals. All message templates emphasize that the student is only seeking career information, as well as explicitly state that the student is not searching for a job. Message templates are in Appendix Figure A1. To summarize, the four message types are:

1. Broad: Asks broadly about the pros and cons of the professional's field.
2. Specific work/life balance: Asks whether work/life balance is a concern in the professional's field.
3. Specific competitive culture: Asks whether competitive culture is a concern in the professional's field.
4. Factual (law only): Asks about the billable hours requirements for a lawyer at a large law firm.

For data science, management consulting, and finance professionals, students sent half of the messages using the broad question, and one quarter of the messages using each specific question. For law professionals, each students sent 44 percent of the messages using the broad question, 22 percent using each specific question, and 4 percent using the factual question. Within each field, professionals were randomly assigned a message type.

In order to estimate the causal effect of student gender on career information received, we randomized whether a professional was sent a message from a male or a female student as well as the specific message type. The random assignment ensures that student characteristics are orthogonal to professional characteristics.

The students sent the messages on weekdays during typical working hours.[5] When a message is sent to a professional, depending on the professional's site preferences, they receive an email notification, an app notification, and/or an alert on the website. After a few days, the site automatically generates a reminder email notification of the message if the professional has not yet responded to the request.

Students were asked to provide the initial responses they receive within 21 days of sending the messages.[6] If a professional responded, the student could choose whether he or she would like to continue the interaction. We emphasized to students that we would not ask for detailed information on these follow-up interactions. As an indication that we selected students based on their genuine interest in career advice, 34 percent of students reported that they planned to stay in touch with at least one of the professionals who responded. Students were asked to not use the site for activities unrelated to the study for the three-week period. We independently verified that students did not change their profile or otherwise engage in site activity

[5] In some cases, students were unable to send all 100 messages in one sitting. In these situations, we asked that the students send the messages as soon as they were able to do so. We recorded the actual date and time that each message was sent.

[6] The vast majority of responses are received within two weeks of sending a message.

throughout the study period. Three weeks after sending the messages, we followed up with the students to ensure that we had received all of their initial responses.

To assess the role of information received on students' future career choices, three weeks after sending messages, students filled out a survey with their career intentions. The survey also asked whether students had follow-up interactions with professionals.

Methodological Advance and Identification

In several ways, our experimental design resembles a traditional correspondence study, in which researchers send fictitious resumes to employers in order to estimate the causal effect of job applicant characteristics on callback rates. In such a design, the format of the resume, the information provided in the resume, and other aspects of the correspondence are controlled by the researcher. The advantages of creating fictitious applicants are numerous: there is precise control over attributes, the researcher avoids dealing with the complexities of the characteristics and behaviors of real people, the study is logistically straightforward to implement, the study is generally low cost to researchers, and by design the applicant characteristic of interest is orthogonal to other applicant characteristics as well as to employer characteristics (Pager, 2007; Bertrand and Duflo, 2017).

In our study, we similarly maintain precise control over the text of the messages sent to professionals, and student characteristics are orthogonal to professional characteristics. In contrast to a traditional correspondence study, we incorporate real students who are interested in information on careers. By incorporating real individuals who are genuinely interested in the outcome of the interaction, this paper is able to address the frequently cited ethical consideration of correspondence and audit studies that research subjects' (i.e., professionals' or employers') time is being wasted (Pager, 2007; Bertrand and Duflo, 2017). The potential loss of time associated with reviewing a job application or providing career information could be substantial. To our knowledge, this is the first study to propose and implement the methodological advance to correspondence studies of incorporating individuals genuinely interested in the interaction in order to resolve this issue.

The incorporation of real students poses some challenges with regard to identification of the causal effect of student gender, however. We cede control over the attributes of students. In particular, we cannot ensure that other student characteristics are orthogonal to student gender. It is likely that student characteristics such as college major will be correlated with student gender. This means that the effect of student gender on the responses that student receive could be confounded by other student attributes.

The online setting serves to mitigate concerns that other student characteristics confound the effect of

student gender. As discussed above, we ask students to limit the information provided on their profiles. But even on characteristics visible on the profile – such as their college majors, graduation year, and number of connections on the platform – male and female students may differ. Our regression specifications control for all student characteristics that are directly observed on the site. There still remains the possibility that there is publicly available online information on the student participants that would, for example, show up if a professional chose to conduct an online search of a student's name and university affiliation. This additional student information could influence the professionals' propensity to respond and the information that they impart. To address this concern, we test whether the effect of student gender is sensitive to the inclusion of student characteristics from the background survey and whether the student has an online presence aside from their profile. We also examine whether the results are robust to restricting the sample to students without an online presence.

3 Data, Sample Restrictions, Summary Statistics

Data

We collect data on response rates, the text of initial responses, and follow-up interactions with professionals. We analyze the text of initial responses using manual classification, sentiment analysis, and natural language processing tools that characterize word distributions. For responses to the broad question, manual classification entails coding whether the response mentions work/life balance or competitive culture. We code mentions of work/life balance using the following definition from the Cambridge dictionary "the amount of time you spend doing your job compared with the amount of time you spend with your family and doing things you enjoy."[7] This includes explicit references to work/life balance, as well as discussion of the hours worked per week, extent of work-related travel, and conflict between/accommodation of work responsibilities and other life priorities. Paraphrased examples of these mentions include:

> [Management consulting] Management Consulting can be considered a lifestyle since it requires travel, very long hours, always being on, and client-specific knowledge.

> [Law] Cons - competitive job market, too much debt from law school, and unpredictable work schedule at times.

We code mentions of competitive culture when the response explicitly mentions competition within the workplace or among coworkers. Due to the low frequency of mentions of competitive culture, we also create

[7] https://dictionary.cambridge.org/us/dictionary/english/work-life-balance

a broader metric of workplace culture, which includes descriptions of interpersonal relations among colleagues, the work environment, or ethical issues in the workplace. Paraphrased examples of these mentions include:

> [Finance] Though this is changing, finance sometimes still depends on connections, bribes, or corruption.

> [Management consulting] Management Consulting is very competitive. Prestigious firms often have an "up or out" policy and your colleagues will fight aggressively to get promoted.

In order to analyze whether mentions of work/life balance or workplace culture crowd out other information, we manually code each component of the response. Specifically, we construct a rubric based on the O*NET classification of occupations' work contexts and activities. Our rubric supplements the O*NET classification with additional fields that are mentioned in the message (such as job search advice or compensation), to ensure we categorize the vast majority of the message text. Since many O*NET categories are used infrequently, we group related categories together. For example, we group interpersonal career attributes such as "communicate with persons outside the organization," "communicate with supervisors, peers, or subordinates," and "deal with external customers." The Online Appendix provides information on each grouping used, as well as the remaining ungrouped but frequently used categories.[8] This rubric allows us to test whether the responses to male and female students exhibit other differences in content.

For the responses to the specific questions, we manually classify whether the response confirms that work/life balance or culture is a concern, refutes that it is a concern, or says "it depends" on various factors like the company, more granular occupation, etc. For the responses to the factual questions, we manually extract the answer to the billable hours requirement, which is a numerical value of hours or numerical range of hours. For answers with a range of hours, we take the midpoint of the range.

For all responses, we hire undergraduates (who are not experiment participants) to provide their subjective evaluations of the tone of the responses, specifically, whether the response would cause a typical undergraduate student to be more or less concerned about work/life balance or workplace culture in the professional's field. We also collect data from the study participants on their impressions of the informativeness and tone of the responses. The data on follow-up interactions is used as an additional indicator of the quality of the professional-student interaction.

To analyze the role of professional attributes in generating gender differences in information received, we collect publicly available information on professionals on this site, including their educational background, gender, and network thickness. We use profile pictures and textual information to assign the gender of each professional. In cases where a picture or text-based information on gender was not available on their profile,

[8] We consider frequent usage to be attributes that appear in more than 5 percent of messages.

we assign gender based on the professional's first name using U.S. Census and Social Security Administration name files. This process successfully classified gender for 99.5 percent of professionals.

Sample Restrictions

The study recruited 100 college students to send messages to approximately 10,000 professionals. One student (and 100 professionals) was used for a pilot and is excluded from the analysis. Five students withdrew due to unforeseen logistical issues with their profiles or with sending the messages. Of the 94 students who were able to successfully create a profile and send messages, 89 students provided data on the responses they received. The five students who dropped out after sending messages constitute sample attrition. We diligently followed up with all student participants, and found that students who took a very long time to send responses (>4 months) had similar response rates to those students who completed the study promptly. This fact makes us less concerned that students who dropped out or who did not reply after sending messages did so because of the replies they received.

We limit the sample to students whose first names unambiguously convey their true gender. We do so using the U.S. Census and Social Security Administration name files. If a student's name is at least 90 percent male or female, and coincides with the student's actual gender, then the student is included in the main analysis. This sample restriction drops 13 students. Our final sample for the analysis consists of 76 students who contacted 7,602 professionals, across four career categories.[9]

Summary Statistics

Summary statistics for the students in the final sample are reported in Table 1, overall and by student gender. The top panel presents student attributes that are visible on or can be easily inferred from the student's profile. Among all students, 58 percent are female. The students are primarily freshmen and sophomores, and 62 percent are economics majors. The substantial representation of economics majors is consistent with our recruiting strategy and the fact that the four career paths chosen are those that economics majors primarily enter post-graduation.

Many students already had profiles on this platform, which is reflected in their number of network connections. Other students – 14 percent – created a profile for the first time through this study. We also record whether students had any information on their profile beyond what is listed in the top panel of this table or had another profile issue that precluded perfect compliance with the profile restrictions. In general, these profile issues were limited to minor deviations from protocol such as a few activities or skills being

[9] Our final sample of professionals was 10,003, so three students were assigned 101 professionals. Two of these students are in the final student sample.

visible on the profile. We control for the deviations from profile restrictions in all regressions.

In a background survey that students filled out prior to sending messages, we collected information on student attributes that are partially observed based on profile information, may be found elsewhere online, or correlated with information found online. For example, student race/ethnicity may be inferred from students' names and first generation college goer could be correlated with the extracurricular activities students are involved in (Jack, 2019). Students are evenly split between race/ethnicity categories and 22 percent are first generation college-goers. The majority of students have some online presence aside from their profile on this site. While male and female students are overall similar, we observe that female students are less likely to be economics or STEM majors, have fewer network connections, and are more likely to identify as Asian/Asian-American.

Summary statistics for the professionals are found in Table 2, overall and by career path. One third of professionals are female, and this varies substantially across field, with representation the lowest in finance and the highest in law. The professionals are, on average, in their late 30s. Professionals were selected based on their attendance of a top 40 U.S. News and World Report university for some part of their education and this is reflected in the selectivity of undergraduate institutions and the substantial fraction who attended an Ivy League university. More than 20 percent of the professionals are alumni of the student's college, with a lower fraction (15 percent) among lawyers. The majority have well-established networks on this site.

Appendix Table A1 presents summary statistics for the main outcomes, including response rates, response length and mentions of work/life balance and workplace culture in responses to the broad question. The overall response rate across all question types is 12 percent, with a lower rate of response to the broad question (10 percent) and the highest response rates to the specific work/life balance and competitive culture questions (14 and 15 percent, respectively).[10] In Appendix Figure A2, we observe that the distribution of responses is centered around 12. Responses are 435 characters, on average, with longer responses to the specific work/life balance question and shorter responses to the factual question. Among responses to the broad question that asks about the pros/cons of the professional's field, 11 percent bring up work/life balance issues and 12 percent mention workplace culture.

Balance Tests

Appendix Table A2 reports results from tests of covariate balance. Each entry in the table is the coefficient on student female in a regression of a professional characteristic (e.g. whether the professional has work experience in data science, the college graduation year of professionals) on an indicator for whether the

[10]This response rate is higher than a correspondence study that sends pitch emails to venture capitalists but lower than studies that send emails to politicians or academics (Gornall and Strebulaev, 2019; Kalla et al., 2018; Milkman et al., 2015).

student who contacted the professional is female. Professional characteristics are balanced across students, indicating that the randomization was successful.

4 The Effect of Student Gender on Information Received

4.1 Econometric Framework

In order to estimate the effect of student gender on information received, we use the following regression specification:

$$Y_{im} = \alpha + \beta StudentFemale_i + \gamma X_{im} + \epsilon_{im} \tag{1}$$

where the dependent variable, Y_{im}, is an outcome such as an indicator for whether a message m sent by student i receives a response, or whether the response that student i receives to message m mentions a specific career attribute. The independent variables are an indicator for whether the student is female, $StudentFemale_i$, as well as a vector of message and student controls, X_{im}. In our baseline specification, we include controls for message characteristics: categorical variables for the day of the week and the time of day that the message was sent, a linear term for the date that the message was sent, and the field of the professional to whom the message was the sent. We also include controls for student characteristics that are directly observable on the site: college major (economics, STEM, other), expected college graduation year, number of network connections, and whether the student was completely compliant with the profile restrictions.[11] Standard errors are clustered at the student level.

While the online setting limits the concern that student characteristics aside from those directly observable on the site confound the effect of student gender on response rate, it is possible that professionals find additional information on students through an online search. We test whether the coefficient on student gender is sensitive to the inclusion of additional student characteristics that may be available elsewhere online. These additional controls include student race/ethnicity, college GPA, first generation student status, and an indicator for whether there is information publicly available on the student through an online search. College GPA and first generation student status are known to be correlated with the extracurricular activities students are involved in, which may be visible online (Jack, 2019). Since student race/ethnicity could also be conveyed through the student's name, we consider this variable partially observed based on information in the student's profile. In the regression results below, we estimate a separate specification to test sensitivity of the main results to this specific control. As an additional check on whether students' online presence

[11] Some students who had profiles prior to the experiment were unable to completely remove all information from their profile. This extra information may include site activity, relevant labor market skills, and extracurricular activities.

confounds the results, in the Appendix we limit the sample to students with no online presence aside from their profile on the professional platform.

4.2 Broad Question

Response Rates

We start by testing whether student gender affects response rates to the broad question that asks about the pros/cons of the professional's field. We estimate equation (1), and use as the dependent variable an indicator for whether a message received a response from the professional. The results are reported in Table 3 columns 1 and 2. Column 1 presents the results with the baseline message and student controls. We observe that response rates to male and female students are very similar; the coefficient on $StudentFemale_i$ is 0.011 and statistically insignificant. Consistent with the notion that the effect of student gender is not confounded by other student characteristics, when we include the supplemental student characteristics that may be observable elsewhere online, the coefficient on student female exhibits little change (column 3). Based on these results, we conclude that professionals are just as willing to engage with male and female students who sent the broad message.

Work/Life Balance

Given that response rates to male and female students are similar, we next analyze whether there are gender differences in the content of the responses to the broad question. As mentioned in the Introduction, we focus on two career attributes that are known to differentially affect the occupational and job choices of women relative to men: work/life balance and competitive culture. We restrict the sample to responses received, estimate equation (1), and use as the dependent variable an indicator for whether the response mentions a work/life balance issue, including work hours, travel, lifestyle, or family/personal life considerations. Table 4 reports the results. Overall, 11 percent of responses bring up work/life balance issues (Appendix Table A1). We observe, however, that responses to female students mention these issues more than twice as much as responses to male students. Among responses to male students, 6.7 percent mention a work/life balance issues. Using the estimates in column 1, the rate for female students is 8.7 percentage points higher. When controls for student race/ethnicity are included in column 2, the coefficient on student female declines slightly to 0.072, indicating that student race/ethnicity could be conveyed through student names and is correlated with student gender. The further inclusion of student characteristics potentially observable elsewhere online marginally increases the coefficient to 0.076.

In Appendix Table A3 we investigate whether the additional information provided on work/life balance

to female students is driven by two O*NET work context categories and one supplemental category: (1) the duration of a typical workweek, (2) flexibility of work schedules, and (3) the ability (or inability) to work from home. We find that there are pronounced gender differences in mentions of the duration of a typical workweek. Responses to female students are 5.4 percentage points more likely to hear about the duration of the typical workweek relative to male students. Given that the male mean for this category is 4.7 percent, female students hear about weekly hours at more than twice the rate of male students. It also appears that responses to female students are more likely to contain information about work schedule flexibility, but this gender difference is not statistically significant.

Mentions of work/life balance tend to be negative. Below are paraphrased examples of responses that mention work/life balance:

> [Law] A career the law opens many doors...and also offers long hours, hard work, firm deadlines, and many challenges.

> [Finance] Challenges can be the hours depending on the area of finance (corporate finance FPA, consulting, investment banking, or even accounting).

Using subjective evaluations from a team of college students who were not study participants, we characterize the anticipated effect of the responses. In particular, we ask the students to rate the extent to which a response would make a typical college student more or less concerned about work/life balance (workplace culture) in the professional's field. Based on the students' evaluations, responses containing mentions of work/life balance increased concern about this issue more than 75% of the time. Only three percent of such responses made students less concerned about work/life balance. Consistent with our finding that female students receive more information on work/life balance, when we consider all responses, we find that responses to female students are more likely to increase concern about work/life balance, but this contrast is not statistically significant (see Appendix Table A4).

Workplace Culture

We also estimate gender differences in mentions of workplace culture. Although the question that asks specifically whether workplace culture is a concern focuses on competitive culture, only six of the responses to the broad question explicitly mention competitive aspects of workplace culture. Due to this low rate, we instead test for gender differences in mentions of workplace culture more generally, which includes descriptions of interpersonal relations among colleagues, the work environment, or ethical issues in the workplace. Twelve percent of all responses to the broad question mention workplace culture (Appendix Table A1). Table 4

columns 3 and 4 report the results from estimating equation (1), which tests whether responses to female students include more mentions of workplace culture. In column 3, the specification with the basic controls for message characteristics and student characteristics available on the profile, the point estimate for the coefficient on student female is close to zero. The coefficient declines a bit and remain statistically insignificant with the inclusion of additional student controls.

We additionally test for mentions of work/life balance and workplace culture, accounting for message non-response in Appendix Table A5. If a student does not receive a response to a message, then it is coded as not containing a mention of work/life balance. The results for both work/life balance and workplace culture are similarly differentiated by student gender.

4.3 Specific Questions

One reason the professionals may differentially emphasize work/life balance issues to female students relative to male students is that the professionals statistically discriminate, that is, believe female students care more about these issues. In this section, we investigate whether gender differences in responses persist when students specifically ask whether work/life balance is a concern in the professional's field.

Response Rates

In Table 3 columns 4-6 we find that, in contrast to the broad question, student gender does affect professionals' propensity respond to the work/life balance question. Considering the baseline specification from column 4, female students are 3.7 percentage points, or 26 percent, more likely to receive a response relative to male students. Again this estimate is insensitive to the inclusion of additional student covariates. This means even when male students specifically request information on work/life balance, they obtain less of it. Consistent with the result that workplace culture is not differentially emphasized to female students in the broad question, there is no gender difference in response rates to the specific question on competitive culture (columns 7-9). This result is not driven by professionals' unwillingness to engage with students on this topic; in fact, the specific competitive culture question had the highest response rate.

Content

The text of the specific questions is designed to elicit a "yes" or "no" response. The message first describes that the student has heard work/life balance (cutthroat culture) is a challenge in the professional's field. Then the message asks directly whether the professional thinks this is a valid concern. Among the responses that answer the question in the text of the message, we manually classify the responses into (1) those that

say yes, it is a concern, (2) those that say it depends on the workplace or specific career path in the field, and (3) those that say no, it is not a concern.

Summary statistics from Appendix Table A1 show that overall the responses to the work/life balance question confirm that work/life is a concern in the professional's field. Only seven percent state that work/life balance is not a valid concern. Approximately half state that work/life balance depends on the firm, specific path, boss, etc., and 44 percent report that it is indeed a concern. In contrast, 30 percent of the responses to the competitive culture question state that it is not a concern in the professional's field. Only 16 percent confirm it is a concern, and the remaining 54 percent state it depends.

Given that responses to the work/life balance question overwhelmingly validate the concern, it is possible that these responses could make college students more concerned about work/life balance in the professional's field. Two paraphrased examples of responses are below:

> [Law] It's definitely a valid concern. At a large law firm, your schedule will be outside of your control. You will not have your evenings, weekends or vacations. In-house is usually better in terms of weekends and vacations, but it is still very demanding.

> [Management Consulting] Yes, would expect between 60-80 hours of work per week and little predictability Mon-Thurs on hours. Weekends are usually open though.

Based on the evaluations of undergraduate students who were not study participants, more than half of the responses to the work/life balance question make students more concerned about work/life balance in the professional's field.

The results demonstrate that women receive more responses to the work/life balance question, and that overall these responses tend to make students more concerned about work/life balance in the professional's field. We additionally test whether there are gender differences in the content and tone of the responses to the specific questions. The results are reported in Appendix Tables A4, A6, and A7. While the coefficients are imprecisely estimated, the responses to female students do not display meaningful content or tone differences relative to the responses to male students.

4.4 Factual Question

The specific questions allow us to rule out the possibility that professionals differentially highlight work/life balance to female students in responses to the broad question because they believe female students care more about this issue. Answers to the specific work/life balance question may still differ for male and female students because the true answer to the question is gender-specific—women may struggle more on average

with work/life balance—and professionals simply report this gender-specific answer. Another possibility is that professionals deem the topic to be more important to female students and are more willing to discuss this topic with them. In the factual question, we shut down the possibility the true answer is gender-specific. Recall that the factual question asks about the minimum billable hours requirement for a lawyer at a large law firm. The minimum billable hours requirement question has an objective answer and the answer should be invariant to the gender of the lawyer. If professionals nonetheless give different responses about a fact to students based on gender, then at least part of the gender difference in emphasis on work/life balance is due to professionals' determination of which topics are more important for male vs. female students.

We find that professionals are 80% more likely to respond to female students than male student asking the factual question (Table 3 columns 10-12), though the coefficient is not consistently statistically significant. Since the true answer to this question is not gender specific, this suggests that professionals' responses are motivated by a subjective determination of what is important for male and female students to know. The majority of responses include a numeric value or range, for example, 1,900 or 2,000-2,100 hours per year. We extract the numeric answers and either use the exact value or, if a range of values is provided, the midpoint of the range. Interestingly, on average, students are quoted an average hours requirement of 1,989, which exceeds the average amount that is publicly listed on the National Association for Law Placement (NALP) online directory, which is 1,921.[12] Appendix Table A8 tests for gender differences in hours quoted to students. While the point estimates suggest that women are quoted higher hours requirements, unfortunately, we are underpowered to detect large differences in hours quoted to male and female students. We based on our power calculations on the results of a small pilot, which showed both substantially higher response rates to this question and also substantially smaller variation in the numeric replies. Although the results are inconclusive due to lack of statistical power, we find it noteworthy that professionals give such variable responses to a question that has an objective answer, which can be found with publicly available data.

To summarize the main results, we find that professionals differentially emphasize work/life balance to female students. This emphasis is apparent both in responses to the question that asks broadly about the pros and cons of the professional's field, as well as in higher response rates to the specific work/life balance question, which generally confirm that work/life balance is a concern. In contrast, we find that professionals do not differentially emphasize workplace culture to female students.

[12]One can access the online directory here: `https://www.nalpdirectory.com/`. We conducted a search of law firms exceeding 1000 employees in the students' city.

5 Implications of Gender Differences in Information Received

We probe the implications of the differential provision of work/life balance information to female students. Specifically, we ask whether work/life balance information crowds out other (potentially useful) information regarding careers and leads to divergence in the career plans of male and female students.

Crowd-out: Does the additional information on work/life balance that women receive crowd out other, potentially useful, information on careers? We investigate this question in two ways. First, we focus on gender differences in response length. Appendix Table A9 reports response length differences across all questions, parameterizing length using character count and the natural logarithm of character count. There are no significant gender differences in the length of the replies. This suggests that the additional emphasis on work/life balance to female students may displace other information.

In order to investigate gender differences in other response components, we comprehensively categorize the text of each response using manual classification, as discussed in Section 3. A description of each frequently used category is in the Online Appendix. Appendix Table A10 reports the results. Responses to female students are less likely to offer any type of advice and less likely to state the professional's qualifications for answering the question. Responses to female students are also less likely to explain career paths and provide information on how to find a job, but these differences are not statistically significant. Responses to female students are more likely to discuss the analytical aspects of a career, compensation, and qualities of individuals who like/succeed in the field, as well as provide an offer to discuss further, but again these contrasts are not statistically significant. A joint test of significance indicates that we can reject that the gender differences are jointly zero. Combining the results on response length with the gender differences in other response content, we find evidence consistent with work/life balance crowding out other career information.

Career choice: To investigate the ramifications of gender differences in information for students' career choices, we analyze students' career plans, which were reported in a survey administered at the conclusion of the study. Of the 76 students in the main sample, 73 completed the survey. We regress an indicator for whether a student was dissuaded from their preferred career path on student characteristics and outcomes from the experiment:[13]

$$D_i = \gamma_0 + \gamma_1 StudentFemale_i + \gamma_2 M_i + \gamma_3 X_i + \varepsilon_i$$

where M_i include characteristics of the messages received by student i, such as whether the student received

[13]Students were asked "Relative to when you began sending messages for this study, are you, on a scale of 1-10, much less likely (1) ... much more likely (10) to go into [data science/finance/law/management consulting]?" We measure dissuasion, D_i, as an indicator for responses of 4 or below to this question for student i's preferred career path.

responses concerning work/life balance, and X_i is a vector of student characteristics, including the basic student controls, race, and student's preferred career path. We note that this analysis is descriptive. It is possible that female students update their career plans over the three-week study period differently than male students, for reasons unrelated to the study itself. That said, we view the following results as suggestive evidence of the role of information in generating gender gaps in career choices.

Relative to the start of the study, women are more dissuaded from their preferred career path than are men. Table 5 column 1 shows that female students are 9.5 percentage points more likely to be dissuaded from their preferred career path. When we add controls for other student characteristics as well as students' preferred career path, the gender difference increases to 11.4 percentage points. Next we test whether the greater emphasis on work/life balance to female students explains this gender difference in career plans. Table 5 column 3 includes controls for whether a student received a response to the broad question that mentions work/life balance and whether the student received a response to the specific work/life balance question. The inclusion of these controls reduces the coefficient on student female by nearly a half, indicating that part of the gender difference in career plans can be attributed to the different information female students received concerning work/life balance. In contrast, the female coefficient is unchanged when we control for mentions of workplace culture or whether the student received a response to the specific workplace culture question (column 4). In column 5, we include controls for work/life balance and workplace culture as well as for message length, and find that message characteristics beyond emphasis on work/life balance do not explain more of the gender gap in career plans.

Women continue to be underrepresented in the four fields that we study. A common way to address the under-representation of minorities in educational and occupational settings is to provide them with accurate information or encouragement (Bursztyn et al., 2020; Porter and Serra, 2020; Bayer and Rouse, 2016). Our results suggest that the information transmitted through informal exchanges served to discourage women from entering their preferred career path, potentially dampening the effects of initiatives intended to increase female entry.

6 Mechanisms and Additional Results

We test several mechanisms that could generate the differential provision of information to male and female students, including the composition of professionals who respond and the differential treatment of students by certain subgroups of professionals.

Composition of professionals who respond: To investigate whether differences in the types of professionals who respond to male and female students contribute to gender differences in responses, we modify

equation (1) as follows:

$$Y_{imp} = \alpha + \beta StudentFemale_i + \gamma X_{im} + \delta X_p + \epsilon_{imp} \quad (2)$$

where the dependent variable is defined as above, but now we introduce the subscript p, which represents that a student contacts or receives a response from professional p. $StudentFemale_i$ and X_{im} are defined as above. We test the sensitivity of β, the coefficient on $StudentFemale_i$, to the inclusion of controls for professionals' characteristics. Specifically, we include professional gender, undergraduate graduation year, undergraduate institution selectivity, whether the professional is an alumnus of the student's college, whether the professional has a graduate degree, and the extent of their network thickness. Standard errors are clustered at the student level.

If different types of professionals respond to male and female students, and their willingness to bring up work/life balance issues is correlated with the professionals' characteristics, then we would expect the inclusion of professional controls to attenuate the coefficient on $StudentFemale_i$. This could happen if, for example, female professionals are more likely to respond to female students, and are also more inclined to bring up work/life balance. Tables 6 and 7 report the results of this test for response rates and mentions of work/life balance and workplace culture in responses to the broad question, respectively. For each outcome, the first specification repeats the results from our preferred specification with message, student profile, and student race/ethnicity controls. The second specification additionally includes the professional controls, as outlined in equation (2). Across all outcomes, the coefficient on $StudentFemale_i$ is invariant to the inclusion of professional controls, supporting the notion that differences in the composition of professionals who respond to male and female students do not explain the greater emphasis on work/life balance to women.

Heterogeneity by professional characteristics: Next we investigate whether the differential provision of information to female students is concentrated among certain subgroups of professionals. We conduct heterogeneity analysis based on professionals' characteristics, specifically professional gender, college graduation year (a proxy for age), selectivity of undergraduate university attended, whether the professional is an alumnus of the student's university, and the professional's field. The results are reported in Appendix Table A11, where each entry represents the coefficient on $StudentFemale_i$ from a separate regression, with the subgroup of professionals in the column title. Panels A, B, and C explore heterogeneity in gender differences in response rates, while panels D and E show the results for mentions of work/life balance and workplace culture in the responses to the broad question. Gender differences in response rates to the broad question exhibit little heterogeneity based on professional characteristics. Response rates to the specific question on work/life balance tell a different story. The higher response rates for female students to this question are

concentrated among younger professionals, those who are alumni of the students' college, and professionals in finance and management consulting. In the responses to the broad question, the greater emphasis on work/life balance issues to female students is concentrated among female professionals, those who are older, professionals with a degree from an Ivy League university, and those in finance and law (panel D). Prior work has demonstrated that female students' educational and occupational outcomes are affected by the gender of their teacher, mentor, or role model (Athey et al., 2000; Carrell et al., 2010; Beaman et al., 2012; Porter and Serra, 2020; Canaan and Mouganie, 2019). An empirical challenge in this work is separating whether female students are more responsive to female mentors' treatment, or whether female mentors treat female students differently than do male mentors. In Appendix Table A12, we observe that female professionals exhibit similar patterns in the differential treatment of male and female students, relative to male professionals. These results provide suggestive evidence that the effect of mentor gender on female students are driven by the responsiveness of female students to the gender of the mentor, rather than mentor gender determining the treatment of students.

Natural language processing: Finally, we non-parametrically test whether responses to male and female students differ. We compute the Kullback-Leibler (K-L) divergence metric, which assesses whether the distributions of words used in responses to male and female students are different. This metric is widely used in the computational linguistics literature and has been previously used in the economics literature by Bohren et al. (2018). We find no significant differences in the distribution of words in responses to male and female students. In order to test whether there are gender differences in the tone of responses to students, we use the NRC lexicon, which contains 10 separate sentiments, and the Bing lexicon, which contains positive and negative sentiments. We find no significant differences in the tone of the responses to male and female students. The Online Appendix describes these results in detail.

7 The Role of Selection

In order to estimate average gender differences in information received, we randomly assigned professionals to students. We find that, on average, professionals differentially emphasize work/life balance issues to female students. However, if in reality female students seek out different professionals than male students, average differences may not be informative of students' experiences. For example, in the labor market, the extent to which minorities apply to discriminatory firms determines wage gaps, not the bias of the average employer (Becker, 1971; Heckman, 1998; Charles and Guryan, 2008).[14] It is reasonable to believe that individuals

[14]In a search framework, average discrimination may generate wage gaps (Black, 1995), but the magnitude of such differences depends on both supply-side and demand-side behavior.

do not sample randomly from their full choice set. For example, Pager and Pedulla (2015) documents that minority job seekers search more broadly for jobs than their non-minority counterparts. Abel (2017) further shows that immigrant job seekers are more likely to search for jobs farther away if they live in areas with higher levels of discrimination. Using a large-scale correspondence study, Agan and Starr (2020) find that Black-sounding job applicants experience more discrimination in less Black neighborhoods, and simulate how equilibrium racial discrimination is affected by residential sorting.

In our setting, if female students prefer to gather information from younger professionals, and only older professionals bring up work/life balance concerns, then our results on average differences in the rate of mentions of work/life balance may not reflect experienced differences. In this section, we use information on student preferences to assess whether student selection of professionals amplifies or attenuates average gender differences in information received. To our knowledge, this is the first paper to tackle the question of whether average differences estimated in an audit or correspondence study reflect differences incorporating agent selection.

7.1 Potential Outcomes Framework for Incorporating Selection

We outline a potential outcomes framework that incorporates student selection into estimates of average gender differences. Let $y_p(1)$ be the response of professional p to a female student and $y_p(0)$ be the response of the same professional to an equivalent male student, asking the same question. Our main experiment described in Section 2 allows us to estimate β, the average gender difference in the responses of professionals in our sample:

$$\beta = E\left(y_p(1) - y_p(0) | p \in 1, ..., P\right)$$

However, students may prefer some professionals over others. Let the set of professionals preferred by students be $\mathbf{P} \subseteq \{1, ..., P\}$. This preferred set of professionals may be younger than than the average in the full sample or may have attended more selective universities. Given information about student preferences over professionals, we can define β^p to be the average gender difference in responses among the preferred professionals:

$$\beta^p = E\left(y_p(1) - y_p(0) | p \in \mathbf{P}\right)$$

To the extent that female students have different preferences over professionals than male students, β^p may not reflect the gender bias experienced by female students. In particular, female students may seek out less biased (or more biased) professionals. Let

$$\beta^m = E\left(y_p(1) - y_p(0) | p \in \mathbf{P^M}\right)$$

and

$$\beta^f = E\left(y_p(1) - y_p(0) | p \in \mathbf{P^F}\right)$$

where $\mathbf{P^M}$ is the set of professionals preferred by male students and $\mathbf{P^F}$ is the set of professionals preferred by female students. β^m is the average gender bias of professionals preferred by male students, and β^f is the average gender bias of professionals preferred by female students.

Even if $\beta^f = \beta^m$, professionals preferred by female students may have different levels of y than the professionals preferred by male students. For example, the rate of response to a message about work/life balance may be lower, on average, for professionals selected by female students. We define the average responses to male students in the set of professionals preferred by males and females, respectively, as:

$$\alpha^m = E\left(y_p(0) | p \in \mathbf{P^M}\right)$$

and

$$\alpha^f = E\left(y_p(0) | p \in \mathbf{P^F}\right)$$

Together, these moments can be used to characterize β^s, which represents how responses to female students seeking information from female-preferred professionals differ from responses to male students seeking information from male-preferred professionals:

$$\beta^s = \underbrace{\beta^f + \alpha^f}_{E(y_p(1)|p \in \mathbf{P^f})} - \underbrace{\alpha^m}_{E(y_p(0)|p \in \mathbf{P^m})}$$

Our experimental design paired with information on student preferences allow us to estimate average gender bias among the preferred professionals, β^p; average gender bias among the professionals preferred by female students, β^f; average gender bias among the professionals preferred by male students, β^m; experienced bias, β^s; and the levels of average response, among male- and female-preferred professionals, α^m and α^f.

We note here that $E\left(y_p(1) | p \in \mathbf{P^f}\right)$ and $E\left(y_p(0) | p \in \mathbf{P^m}\right)$ are available in observational data, meaning that one can estimate $\beta^s = \beta^f + \alpha^f - \alpha^m$. However, using only observational data on student selection of professionals and the outcomes of these interactions, the average bias of professionals contacted by male and female students (β^m and β^f, respectively) are not identified unless $\mathbf{P^M} = \mathbf{P^F}$. To see the importance of identifying β^m and β^f, suppose β^s is negative, that is, we observe that female students receive lower response rates than male students, when students contact professionals on their own. From β^s alone, we cannot discern whether steering female students to male-preferred professionals would improve gender disparities in response rates. If male-preferred professionals were very gender biased ($\beta^m << 0$), steering female students toward

these professionals could even exacerbate gender differences in response rates.

In the next section, we discuss our method for identifying preferred professionals. Our method eliminates the correlation between the gender of students who prefer professionals and the gender of students who actually contacted these professionals. We predict student preferences over the entire set of professionals regardless of whether the student actually sent a message to the professionals. Because we observe responses of professionals to a random sub-sample of students, some male and some female, we are able to directly estimate β^m and β^f.

7.2 Estimating Student Preferences for Professionals

To estimate students' preferences over professionals, we asked students to rank professionals in terms of whom they would most prefer to ask the questions in the study. Before sending any messages, students were asked to spend 20 minutes studying the profiles of professionals they would be messaging and rank

1. The five professionals they would be most interested in asking about the pros and cons of the professional's field.

2. The five professionals they would be most interested in asking about work/life balance in the professional's field.

3. The five professionals they would be most interested in asking about workplace culture in the professional's field.

Note that these questions parallel the actual messages sent by students.[15] Students were told that these rankings would not impact the messages they send in the study and were purely for the researchers to learn about their preferences over the professionals. Students were also told that they could choose the same five people for each of the three rankings, different people, or a combination thereof — whatever reflected their true preferences.

Using these student rankings, we estimate a rank-ordered logit choice model for student preferences over professional characteristics (Beggs et al., 1981). As in Table 2, we model preferences over professionals' college graduation year, college selectivity, whether the professional has any graduate degree, whether the professional is an alumnus of the student's university, whether the professional has a degree from an Ivy League institution, the size of the professional's network, the professional's gender, and whether the professional is in the same field as the student's preferred career field.

[15] We did not ask students rank the lawyers whom they would want to ask about minimum billable hours requirements—there is no factual question ranking.

We assume that the value of a particular professional p to a student j is given by

$$V_p^j = \beta^j X_p + \varepsilon_p^j \tag{3}$$

where X_p are the characteristics of professional p, β^j is student j's relative valuation of the professional's characteristics, and ε_p^j is an error term. Assuming that errors are independent and follow the type I extreme-value distribution, the probability that an alternative p is valued higher than other alternatives is given by:

$$Pr\left\{V_p^j > \max_{k\in\{1,\ldots,P\},k\neq p}\{V_k^j\}\right\} = \frac{\exp(V_p^j)}{\sum_{k=1}^{P}\exp(V_k^j)} \tag{4}$$

Assuming that student preferences depend only on their observable characteristics and the characteristics of professionals, and that unranked alternatives are not preferred to ranked alternatives, we estimate a likelihood model based on equation (4) separately for each question-specific ranking (broad, specific-work/life balance, specific-culture). We allow preferences to depend on all student characteristics, all professional characteristics, and their interactions.[16] For each student-professional pair in the full sample, we calculate the predicted probability that the professional is ranked first by the student. We use the predicted probabilities to form the preferred professional sets $\mathbf{P}$, $\mathbf{P^m}$, and $\mathbf{P^f}$ from the full set of professionals in our sample. This method allows us to create a data set that simulates an assignment process in which students chose the 100 professionals they would most prefer to contact from the full set of professionals in the sample (half asking a broad question, and a quarter devoted to each specific question).[17]

The characteristics of preferred professionals are presented in Appendix Table A15. The preferred sample differs in a number of ways from the full sample of professionals: they are younger, more likely to be female, and more likely to have a degree from an Ivy League institution. The professionals preferred by female students relative to those preferred by male students are more likely to have a law or management consulting background, are more likely to be female, and are more likely to have a graduate degree. Given the differences between the preferred professionals and the full sample, as well as the differences between the preferred professionals of male and female students, the average gender differences estimated in Section 4 may not

[16] More specifically, for students we use the following characteristics: gender, network size, additional profile information, economics major indicator, STEM major indicator, graduation year, and an indicator for whether the student had any problem with their profile (for example being slow to upload the profile picture). We interact these student characteristics with the following professional characteristics: gender, binned undergraduate graduation year (1980s or earlier, 1990s, 2000s, 2010s+, or no information on graduation year available), connections (binned in low, medium, and high), an indicator for whether the professional is an alumnus of the student's college, undergraduate institution selectivity quartile, whether the professional has any graduate degree, and an indicator for whether the professional has any degree from an Ivy League institution. We also interact indicators for the student's preferred field with indicators of the professional's field.

[17] We limit the sample of professionals to the random subset who were contacted by students in our final sample and who were actually asked the question that parallels the question generating preferences. For example, when selecting the set of 48 professionals that a student would want to ask the broad question, we select the 48 professionals predicted to be that student's most preferred from the subset of professionals who were actually asked the broad question by a student in the main experiment.

reflect gender differences inclusive of student selection.

7.3 Econometric Framework for Incorporating Selection

To estimate average gender differences in response among the preferred professionals, β^p, we use the following specification:

$$Y_p = \alpha + \beta^p FemaleStudentSent_p + \gamma X_p + \epsilon_p \tag{5}$$

where the sample is restricted to the set of professionals $\mathbf{P}$ preferred by students.[18] Y_p is the outcome of a message sent to professional p, for example, an indicator for whether the professional responded to the message. $FemaleStudentSent_p$ indicates whether in the main experiment the professional p actually received a message from a female student or a male student, X_p is a vector of professional-level controls such as industry, the date and time that a professional was sent a message, and other characteristics of the student who sent the message to the professional.

We estimate β^f and β^m, the gender bias of professionals preferred by female and male students, respectively, as well as $\alpha^f - \alpha^m$, the difference in average response among professionals preferred by male and female students, with the following specification:

$$\begin{aligned} Y_{jp} = a + (\alpha^f - \alpha^m) FemaleStudentPreferred_j + \beta^m FemaleStudentSent_p + \\ (\beta^f - \beta^m) FemaleStudentPreferred_j \times FemaleStudentSent_p + cX_p + dX_j + e_{jp} \end{aligned} \tag{6}$$

where Y_{jp} is an outcome of a message sent to professional p who is preferred by student j, $FemaleStudentPreferred_j$ indicates whether the student j who prefers this professional is female, $FemaleStudentSent_p$ indicates whether in the main experiment the professional p actually received a message from a female student, and X_p is defined as above. We include a vector of additional student controls, X_j, that could be correlated with student gender and influence the student's preferences over professionals. Note that student j (almost) never actually contacted professional p in the main experiment. In the simulated data set, the gender of students who selected a particular professional is orthogonal to the gender of students who actually contacted the professional. This means that we observe the response of professionals preferred by student j to messages sent by *both* male and female students. We can therefore identify the gender bias of professionals preferred by students overall, by female students, and by male students (β^p, β^f, and β^m, respectively).

[18] Notice professionals not preferred by either male or female students are not included in this data set (or alternatively, get zero weight). Note also that a professional can be repeated in this sample if multiple students preferred the same professional.

7.4 Gender Differences in Information Received, Incorporating Selection

Tables 8 and 9 present the estimates of equation (5) and equation (6). Column 1 reports estimates of β^p, which measures gender bias among the professionals preferred by students. Column 2 reports estimates of the average differences in response among professionals preferred by female relative to male students ($\alpha^f - \alpha^m$), the gender bias in responses of professionals preferred by male students (β^m), and the relative gender bias of professionals preferred by female students ($\beta^f - \beta^m$). Column 3 adds controls for the characteristics of student j who preferred the professional. In the bottom portion of Tables 8 and 9 there are also estimates of the average outcome for female students sending messages to female-preferred professionals. This estimate is labeled "Female Pref and Female Sent" and is the sum of the coefficients on the three covariates in the regressions. The difference between these rates and the average outcome for male students sending messages to male-preferred professionals is the gender difference that students would experience incorporating selection (β^s).

We find that student selection of professionals does not mitigate gender differences in information received about work/life balance. Recall that in our main experiment, professionals were 7.2 percentage points more likely to mention work/life balance issues in their responses to female students relative to male students. When we incorporate student selection of professionals, the coefficient on $FemaleStudentSent_p$ in column 1 of Table 9 is 0.237. This means that the professionals preferred by students are more likely in their replies to differentially highlight work/life balance issues to female students than professionals on average. In column 2, we observe that, although female students select professionals who bring up work/life balance less overall to male students than the professionals selected by male students (the coefficient on $FemaleStudentPreferred_j$ is -0.030), female students also select professionals who mention these issues slightly more to women than to men (the coefficient on the interaction term is 0.010). These latter two effects are small and somewhat offsetting, and imply that experienced gender differences in the content of responses, β^s, are larger than average gender differences. In column 3, when we include additional controls for characteristics of student j, these results do not change.

Student selection of professionals also has ramifications for response rates. Similar to the main results, in Table 8 column 1 we find that in the broad question the preferred professionals do not respond differently to male and female students. Female students, however, tend to select professionals with lower response rates (in column 2 the coefficient on $FemaleStudentPreferred_j$ is -0.041).[19] In the main experiment, we find that female students receive more responses to the specific work/life balance question. However, in

[19]The lower response rates of female-preferred professionals reduces the gender disparity in hearing about work/life balance unconditional on response. Accounting for non-response, women are still more than three times as likely to hear about work/life balance issues relative to men among preferred professionals, which is a larger effect than in our main experimental results.

the preferred-professional sample (Table 8 column 4), we see that this difference is substantially attenuated. In addition, female-preferred professionals are substantially less likely to respond to the work/life balance question as well as to other question types. While these results are imprecise, if anything, experienced gender differences in response rates are of opposite sign relative to the average gender differences estimated in the main experiment. The pattern is similar for response rates to the specific competitive culture question. Finally, the preferred professionals' responses are less likely to mention workplace culture to female students relative to male students (Table 9 column 7).

This exercise demonstrates that the causal effect of student gender may misrepresent students' experiences when students do not select professionals at random. In the main experiment, professionals were more likely to respond to female students when students asked about work/life balance. Incorporating student preferences reverses this conclusion because female students select professionals who reply less overall. If we wanted to reduce gender disparities in information access, we might consider improving the response rates to female students by steering them to the professionals preferred by male students. Using these estimates, we can predict the outcome of such a policy. In the bottom half of Table 8, we calculate the counterfactual response rate if female students contacted male-preferred professionals (labeled "Male pref and female sent"). We can contrast this with the experienced response rate for female students (labeled "Female pref and female sent"). Across all question types, there is a substantial gap in the experienced and counterfactual response rates for female students, indicating that such a policy would be effective in increasing response rates to female students.

In contrast, for responses to the broad question, the experienced gender gaps in mentions of work/life balance and workplace culture are amplified relative to the average differences from the main experiment. We can again consider a policy to reduce gender disparities in the content of information provided, by steering female students to male-preferred professionals. If female students contacted male-preferred professionals, there would be no change in the gender gap in the frequency of work/life balance mentions. This is due to the fact that male-preferred professionals bring up work/life balance more frequently, but are also slightly less gender-biased in their responses.

8 Conclusion

Information transmission through informal interactions is an everyday, routine occurrence. As of yet, research on this mode of communication has been limited by the inherently unobservable nature of these interactions. This paper provides a window into informal exchanges and additionally sheds light on a subtle form of disparate treatment of individuals based on their demographic characteristics. In a large-scale field

experiment with college students interested in career advice, we estimate the causal effect of student gender on information provided by professionals regarding career paths. The experimental design also advances the correspondence study methodology by incorporating real individuals who are genuinely interested in the interactions being studied. We find that professionals differentially emphasize work/life balance to female students, even when students do not specifically ask about this issue. When students ask specifically about work/life balance, professionals are more willing to engage with female students on this topic. Evidence suggests that gender differences in information received matter: after the study, female students are more dissuaded from their preferred career path and this difference is in part explained by the greater emphasis on work/life balance in the messages they received.

We combine the results of the field experiment with student preferences for professionals in order to determine whether average disparate treatment of male and female students is amplified or attenuated by student selection of whom to contact for information. Male and female students have different preferences for professionals, but these different preferences do not alter the extent of gender bias in professionals' emphasis on work/life balance. Looking beyond our setting, correspondence and audit studies are powerful tools to study average discrimination in the labor market. Incorporating the preferences of real individuals complements estimates of average discrimination and enriches our understanding of the realized experiences of individuals.

If information access depends on an individual's gender, absent knowing the nature of the missing, inaccurate, or emphasized information, it may be difficult for individuals to correct for these disparities. Research demonstrates substantial gender gaps in knowledge of fundamental economic parameters, which are used to inform consumption, financial, and labor market choices (D'Acunto et al., 2019; Dwyer et al., 2002). Our work points to disparate information access as one plausible determinant of these knowledge and behavior gaps.

References

Abel, Martin, "Labor market discrimination and sorting: evidence from South Africa," *World Bank Policy Research Working Paper*, 2017.

Acemoglu, Daron, Kostas Bimpikis, and Asuman Ozdaglar, "Dynamics of information exchange in endogenous social networks," *Theoretical Economics*, 2014, *9* (1), 41–97.

Agan, Amanda Y and Sonja B Starr, "Employer Neighborhoods and Racial Discrimination," Working Paper 28153, National Bureau of Economic Research November 2020.

Athey, Susan, Christopher Avery, and Peter Zemsky, "Mentoring and Diversity," *American Economic Review*, September 2000, *90* (4), 765–786.

Bayer, Amanda and Cecilia Elena Rouse, "Diversity in the Economics Profession: A New Attack on an Old Problem," *Journal of Economic Perspectives*, November 2016, *30* (4), 221–42.

Beaman, L., E. Duflo, R. Pande, and P. Topalova, "Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India," *Science*, 2012, *335* (January), 582–586.

Becker, Gary, *The Economics of Discrimination*, 2 ed., University of Chicago Press, 1971.

Beggs, S., S. Cardell, and J. Hausman, "Assessing the potential demand for electric cars," *Journal of Econometrics*, September 1981, *17* (1), 1–19.

Bertrand, M. and E. Duflo, "Field Experiments on Discrimination," *Handbook of Economic Field Experiments*, 2017, *1*, 309–393.

Bertrand, Marianne, "New Perspectives on Gender," in David Card and Orley Ashenfelter, eds., *David Card and Orley Ashenfelter, eds.*, Vol. 4 of *Handbook of Labor Economics*, Elsevier, 2011, pp. 1543–1590.

Black, Dan A., "Discrimination in an Equilibrium Search Model," *Journal of Labor Economics*, 1995, *13* (2), 309–334.

Blau, Francine D and Lawrence M Kahn, "The Gender Wage Gap: Extent, Trends, and Explanations," *Journal of Economic Literature*, sep 2017, *55* (3), 789–865.

Bohren, Aislinn, Alex Imas, and Michael Rosenberg, "The Language of Discrimination: Using Experimental versus Observational Data," *AEA Papers and Proceedings*, 2018, *108*, 169–74.

Bursztyn, Leonardo, Alessandra L. GonzÃ¡lez, and David Yanagizawa-Drott, "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia," *American Economic Review*, 2020, (Forthcoming).

Canaan, Serena and Pierre Mouganie, "Female Science Advisors and the STEM Gender Gap," *SSRN Electronic Journal*, 2019.

Carrell, Scott E, Marianne E Page, and James E West, "Sex and Science: How professor Gender Perpetuates the Gender Gap," *The Quarterly Journal of Economics*, 2010, *125* (3), 1101–1144.

Cawley, John and Christopher J. Ruhm, *The Economics of Risky Health Behaviors*, Vol. 2, Elsevier B.V., 2011.

Charles, KerwinÂ Kofi and Jonathan Guryan, "Prejudice and Wages: An Empirical Assessment of Beckerâs The Economics of Discrimination," *Journal of Political Economy*, 2008, *116* (5), 773–809.

Coffman, Lucas C, Clayton R Featherstone, and Judd B Kessler, "Can Social Information Affect What Job You Choose and Keep?," *American Economic Journal: Applied Economics*, jan 2017, *9* (1), 96–117.

D'Acunto, Francesco, Ulrike M. Malmendier, Juan Ospina, Michael Weber, and Michael Weber, "Exposure to Daily Price Changes and Inflation Expectations," Technical Report 2019.

Dinkelman, Taryn and A Claudia Martínez, "Investing in schooling in Chile: The role of information about financial aid for higher education," 2014.

Duflo, Esther and Emmanuel Saez, "The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment," *Quarterly Journal of Economics*, 2003, *118* (3), 815–842.

Dwyer, Peggy D., James H. Gilkeson, and John A. List, "Gender differences in revealed risk taking: Evidence from mutual fund investors," *Economics Letters*, 2002, *76* (2), 151–158.

Flory, Jeffrey A, Andreas Leibbrandt, and John A List, "Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job-Entry Decisions," *The Review of Economic Studies*, 2015, *82* (1), 122–155.

Goldin, Claudia, "A Grand Gender Convergence: Its Last Chapter," *American Economic Review*, April 2014, *104* (4), 1091–1119.

Gornall, Will and Ilya A. Strebulaev, "Gender, Race, and Entrepreneurship: A Randomized Field Experiment on Venture Capitalists and Angels," *SSRN Electronic Journal*, 2019.

Heckman, James J., "Detecting Discrimination," *Journal of Economic Perspectives*, 1998, *12* (2), 101–116.

Hoxby, Caroline M and Sarah Turner, "What High-Achieving Low-Income Students Know about College," *American Economic Review*, may 2015, *105* (5), 514–517.

Jack, Anthony Abraham, *The Privileged Poor: How Elite Colleges Are Failing Disadvantaged Students*, Harvard University Press, 2019.

Jensen, Robert, "The (perceived) returns to education and the demand for schooling," *Quarterly Journal of Economics*, 2010, (May).

Kalla, Joshua, Frances Rosenbluth, and Dawn Langan Teele, "Are You My Mentor? A Field Experiment on Gender, Ethnicity, and Political Self-Starters," *The Journal of Politics*, 2018, *80* (1), 337–341.

Milkman, Katherine L., Modupe Akinola, and Dolly Chugh, "What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations.," *Journal of Applied Psychology*, 2015, *100* (6), 1678–1712.

Montgomery, James D., "Social Networks and Labor-Market Outcomes: Toward an Economic Analysis," *The American Economic Review*, 1991, *81* (5), 1408–1418.

Nguyen, Trang, "Information , Role Models and Perceived Returns to Education : Experimental Evidence from Madagascar â," 2008.

Niederle, Muriel and Lise Vesterlund, "Gender and Competition," *Annual Review of Economics*, 2011, *3*, 601–630.

Pager, Devah, "The use of field experiments for studies of employment discrimination: Contributions, critiques, and directions for the future," *Annals of the American Academy of Political and Social Science*, 2007, *609* (1), 104–133.

_ **and David Pedulla**, "Race, Self-Selection, and the Job Search Process," *American Journal of Sociology*, 2015, *120* (4), 1005–54.

Porter, Catherine and Danila Serra, "Gender Differences in the Choice of Major: The Importance of Female Role Models," *American Economic Journal: Applied Economics*, July 2020, *12* (3), 226–54.

Wiswall, Matthew and Basit Zafar, "Determinants of college major choice: Identification using an information experiment," *Review of Economic Studies*, 2013, *82* (2), 791–824.

_ **and** _ , "How do college students respond to public information about earnings?," *Journal of Human Capital*, 2015, *9* (2), 117–169.

_ **and** _ , "Preference for the workplace, investment in human capital, and gender," *Quarterly Journal of Economics*, 2018, *133* (1), 457–507.

Zafar, Basit, "How Do College Students Form Expectations?," *Journal of Labor Economics*, 2011, *29* (2), 301–348.

Figures and Tables

Table 1: Student Summary Statistics

	All Students	Male	Female
Profile Information			
Female	0.58		
	(0.50)		
Expected Graduation Year	2022.24	2022.50	2022.05
	(1.04)	(0.95)	(1.08)
Economics	0.62	0.69	0.57
	(0.49)	(0.47)	(0.50)
STEM	0.22	0.25	0.20
	(0.42)	(0.44)	(0.41)
0-49 Connections	0.46	0.44	0.48
	(0.50)	(0.50)	(0.51)
50-249 Connections	0.28	0.25	0.30
	(0.45)	(0.44)	(0.46)
250+ Connections	0.26	0.31	0.23
	(0.44)	(0.47)	(0.42)
Profile Extra Info	0.47	0.56	0.41
	(0.50)	(0.50)	(0.50)
Profile Issue	0.07	0.03	0.09
	(0.25)	(0.18)	(0.29)
Demographic Information			
White/Caucasian	0.30	0.28	0.32
	(0.46)	(0.46)	(0.47)
Asian/Pacific Islander	0.37	0.31	0.41
	(0.49)	(0.47)	(0.50)
Other Race/Ethnicity	0.33	0.41	0.27
	(0.47)	(0.50)	(0.45)
Additional Student Information			
GPA	3.64	3.62	3.65
	(0.28)	(0.34)	(0.24)
First Generation College Student	0.22	0.25	0.20
	(0.42)	(0.44)	(0.41)
Online Presence	0.71	0.66	0.75
	(0.46)	(0.48)	(0.44)
Observations	76	32	44

Note: This table reports summary statistics for the final student sample, overall and by student gender. Means for each student characteristic are reported, with standard deviations in parentheses.

Table 2: Professionals Summary Statistics

	All Professionals	Data Science	Finance	Law	Mgmt Consulting
Data Science	0.13				
	(0.33)				
Finance	0.28				
	(0.45)				
Law	0.33				
	(0.47)				
Mgmt Consulting	0.26				
	(0.44)				
Female	0.34	0.29	0.23	0.43	0.36
	(0.47)	(0.45)	(0.42)	(0.49)	(0.48)
College Graduation Year	2003.62	2009.55	2003.83	1998.31	2007.45
	(12.00)	(7.59)	(11.95)	(11.84)	(11.16)
College Selectivity - Admit Rate	0.25	0.39	0.25	0.20	0.28
	(0.22)	(0.28)	(0.22)	(0.16)	(0.23)
Alumni of Student's College	0.21	0.25	0.27	0.15	0.21
	(0.41)	(0.44)	(0.44)	(0.35)	(0.41)
Any Graduate Degree	0.70	0.72	0.50	1.00	0.51
	(0.46)	(0.45)	(0.50)	(0.00)	(0.50)
Any Ivy Degree	0.16	0.07	0.16	0.19	0.15
	(0.36)	(0.26)	(0.36)	(0.39)	(0.36)
0-249 Connections	0.11	0.13	0.10	0.15	0.05
	(0.31)	(0.33)	(0.30)	(0.36)	(0.21)
250-499 Connections	0.21	0.24	0.21	0.27	0.13
	(0.41)	(0.42)	(0.41)	(0.45)	(0.33)
500+ Connections	0.64	0.59	0.64	0.54	0.78
	(0.48)	(0.49)	(0.48)	(0.50)	(0.41)
Observations	7602	970	2156	2522	1954

Note: This table reports summary statistics for the sample of professionals, overall and by professional field. Means for each professional characteristic are reported, with standard deviations in parentheses.

Table 3: Effect of Student Gender on Response Rates, By Question Type

	Broad			Work/Life Balance			Competitive Culture			Factual (Law Only)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Student Female	0.011	0.014	0.013	0.037**	0.041***	0.040**	0.003	0.009	0.011	0.060*	0.059	0.060
	(0.011)	(0.011)	(0.010)	(0.015)	(0.015)	(0.016)	(0.015)	(0.017)	(0.017)	(0.032)	(0.036)	(0.039)
Finance	-0.055***	-0.055***	-0.055***	-0.118***	-0.118***	-0.118***	-0.094***	-0.094***	-0.093***			
	(0.018)	(0.018)	(0.018)	(0.037)	(0.037)	(0.037)	(0.032)	(0.032)	(0.032)			
Law	-0.084***	-0.084***	-0.084***	-0.134***	-0.134***	-0.134***	-0.107***	-0.107***	-0.108***			
	(0.017)	(0.017)	(0.017)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)			
Mgmt Consulting	-0.030*	-0.030*	-0.030*	-0.071**	-0.071**	-0.071**	-0.086***	-0.087***	-0.088***			
	(0.018)	(0.018)	(0.017)	(0.032)	(0.032)	(0.032)	(0.031)	(0.031)	(0.031)			
Male Mean	0.101			0.130			0.139			0.073		
Observations	3530	3530	3530	1763	1763	1763	1776	1776	1776	298	298	298
Message Time/Date	X	X	X	X	X	X	X	X	X	X	X	X
Student Profile	X	X	X	X	X	X	X	X	X	X	X	X
Student Race/Ethnicity		X	X		X	X		X	X		X	X
Additional Student			X			X			X			X

Note: This table reports the results of the estimation of the regression specification outlined in equation (1), in which the dependent variable is an indicator for whether a message received a response, and the independent variables are an indicator for whether the student who sent the message is female, the professional's field, message time/date characteristics, and student profile characteristics. Separate regressions are estimated for each question type: broad, specific - work/life balance, specific - competitive culture, and factual. Columns 1, 4, 7, and 10 report results from the baseline specification. Columns 2, 5, 8, and 11 report results from a specification that additionally includes controls for student race/ethnicity. Columns 3, 6, 9, and 12 additionally include controls for student characteristics that may be observable elsewhere online. The omitted field is data science. Standard errors are clustered at the student level and are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Effect of Student Gender on Mentions of Work/Life Balance and Workplace Culture

	Work/Life Balance			Workplace Culture		
	(1)	(2)	(3)	(4)	(5)	(6)
Student Female	0.087***	0.072**	0.076**	-0.003	-0.025	-0.025
	(0.032)	(0.031)	(0.032)	(0.034)	(0.035)	(0.035)
Finance	0.013	0.018	0.012	-0.127**	-0.120**	-0.122**
	(0.028)	(0.029)	(0.029)	(0.055)	(0.055)	(0.055)
Law	0.062	0.062	0.053	-0.172***	-0.170***	-0.172***
	(0.042)	(0.043)	(0.041)	(0.053)	(0.054)	(0.054)
Mgmt Consulting	0.207***	0.209***	0.209***	-0.074	-0.071	-0.073
	(0.051)	(0.052)	(0.051)	(0.060)	(0.060)	(0.061)
Male Mean	0.067			0.128		
Observations	363	363	363	363	363	363
Message Time/Date	X	X	X	X	X	X
Student Profile	X	X	X	X	X	X
Student Race/Ethnicity		X	X		X	X
Additional Student			X			X

Note: This table reports the results of the estimation of the regression specification outlined in equation (1), in which the dependent variable is an indicator for whether a response mentions work/life balance (columns 1-3) or workplace culture (columns 4-6), and the independent variables are an indicator for whether the student who sent the message is female, the professional's field, message time/date characteristics, and student profile characteristics. Columns 1 and 4 report results from the baseline specification. Columns 2 and 5 report results from a specification that additionally includes controls for student race/ethnicity. Columns 3 and 6 additionally include controls for student characteristics that may be observable elsewhere online. The omitted field is data science. Standard errors are clustered at the student level and are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Gender Differences in Career Plans: Relative to the Start of the Study, Is the Student Less Likely to Enter Preferred Career Path?

	(1)	(2)	(3)	(4)	(5)
Student Female	0.095**	0.114*	0.058	0.117*	0.059
	(0.046)	(0.064)	(0.051)	(0.068)	(0.058)
Response Mentioned Work/Life Balance			0.185		0.190
			(0.136)		(0.159)
Received Response to Specific Work/Life Question			0.079*		0.081*
			(0.043)		(0.044)
Response Mentioned Workplace Culture				0.045	0.002
				(0.057)	(0.085)
Received Response to Specific Culture Question				0.010	-0.003
				(0.069)	(0.069)
Male Mean	0.000				
Observations	73	73	73	73	73
Industry Controls		X	X	X	X
Student Controls		X	X	X	X
Message Controls					X

Note: Standard errors are clustered at the student level and are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Role of Professional Composition in Gender Differences in Response Rates

	Broad		Work/Life Balance		Competitive Culture		Factual (Law Only)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Student Female	0.014	0.018	0.041***	0.042***	0.009	0.008	0.059	0.055
	(0.011)	(0.011)	(0.015)	(0.016)	(0.017)	(0.017)	(0.036)	(0.040)
Finance	-0.055***	-0.055***	-0.118***	-0.120***	-0.094***	-0.074**		
	(0.018)	(0.018)	(0.037)	(0.037)	(0.032)	(0.034)		
Law	-0.084***	-0.060***	-0.134***	-0.099**	-0.107***	-0.070*		
	(0.017)	(0.020)	(0.033)	(0.038)	(0.033)	(0.037)		
Mgmt Consulting	-0.030*	-0.032*	-0.071**	-0.079**	-0.087***	-0.076**		
	(0.018)	(0.019)	(0.032)	(0.034)	(0.031)	(0.032)		
Observations	3530	3530	1763	1763	1776	1776	298	298
Message Time/Date	X	X	X	X	X	X	X	X
Student Profile	X	X	X	X	X	X	X	X
Student Race/Ethnicity	X	X	X	X	X	X	X	X
Professional		X		X		X		X

Note: This table reports the results of the estimation of the regression specification outlined in equation (2), in which the dependent variable is an indicator for whether a message received a response, and the independent variables are an indicator for whether the student who sent the message is female, the professional's field, message time/date characteristics, student profile characteristics, and student race/ethnicity. Separate regressions are estimated for each question type: broad, specific - work/life balance, specific - competitive culture, and factual. Columns 1, 3, 5, and 7 report results from the preferred specification. Columns 2, 4, 6, and 8 report results from a specification that additionally includes controls for professional characteristics. The omitted field is data science. Standard errors are clustered at the student level and are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Role of Professional Composition in Gender Differences in Responses to Specific Questions

	Work/Life Balance		Workplace Culture	
	(1)	(2)	(3)	(4)
Student Female	0.072**	0.067**	-0.025	-0.026
	(0.031)	(0.033)	(0.035)	(0.038)
Finance	0.018	-0.007	-0.120**	-0.112*
	(0.029)	(0.037)	(0.055)	(0.063)
Law	0.062	0.017	-0.170***	-0.211**
	(0.043)	(0.062)	(0.054)	(0.081)
Mgmt Consulting	0.209***	0.183***	-0.071	-0.064
	(0.052)	(0.054)	(0.060)	(0.063)
Observations	363	363	363	363
Message Time/Date	X	X	X	X
Student Profile	X	X	X	X
Student Race/Ethnicity	X	X	X	X
Professional		X		X

Note: This table reports the results of the estimation of the regression specification outlined in equation (2), in which the dependent variable is an indicator for whether a response mentions work/life balance (columns 1 and 2) or workplace culture (columns 3 and 4), and the independent variables are an indicator for whether the student who sent the message is female, the professional's field, message time/date characteristics, student profile characteristics, and student race/ethnicity. Columns 1 and 3 report results from the preferred specification. Columns 2 and 4, 6 report results from a specification that additionally includes controls for professional characteristics. The omitted field is data science. Standard errors are clustered at the student level and are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Effect of Student Gender on Response Rates, Inclusive of Student Selection

	Broad			Work/Life Balance			Competitive Culture		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female Student Sent	-0.001	-0.007	-0.008	0.012	0.014	0.014	-0.009	-0.055	-0.052
	(0.023)	(0.028)	(0.028)	(0.039)	(0.047)	(0.047)	(0.051)	(0.062)	(0.062)
	[0.030]	[0.041]	[0.041]	[0.049]	[0.061]	[0.063]	[0.072]	[0.090]	[0.091]
Female Student Preferred		-0.041*	-0.024		-0.041	-0.048		-0.112**	-0.113**
		(0.024)	(0.025)		(0.041)	(0.040)		(0.053)	(0.054)
		[0.044]	[0.045]		[0.059]	[0.062]		[0.067]	[0.066]
Female Student Pref x Female Student Sent		0.011	0.012		-0.004	-0.004		0.083	0.082
		(0.031)	(0.031)		(0.052)	(0.052)		(0.068)	(0.069)
		[0.051]	[0.050]		[0.064]	[0.065]		[0.088]	[0.090]
Male Pref and Male Sent Mean		0.137			0.150			0.232	
Male Pref and Female Sent ($\beta^m - \alpha^m$)		0.130	0.129		0.164	0.164		0.177	0.179
Female Pref and Female Sent ($\alpha^f + \beta^f$)		0.100	0.117		0.118	0.112		0.147	0.148
P-value for Experienced F-M diff		0.169	0.463		0.504	0.444		0.214	0.229
Observations	3648	3648	3648	1824	1824	1824	1824	1824	1824
Message Time/Date	X	X	X	X	X	X	X	X	X
Sender Profile, Race	X	X	X	X	X	X	X	X	X
Student Profile, Race			X			X			X

Note: This table reports the results of the estimation of the regression specification outlined in equation (6), in which the dependent variable is an indicator for whether a message received a response, and the independent variables are an indicator for whether the student who sent the message is female, an indicator for whether the professional was in the preferred set of a female student, and the interaction of these terms, as well as the professional's field, message time/date characteristics, and student profile characteristics. Columns 2, 5 and 8 report results from the baseline specification. Columns 3, 6, and 9 report results from a specification that additionally includes controls for characteristics available on their profile and race/ethnicity of the student selecting the professionals. Columns 1, 4, and 7 report results from the regression specification outlined in equation (1), which relative to the above omits an indicator for whether the professional was in the preferred set of a female student. The bottom panel of the table provides the mean response rate to male students among professionals preferred by male students, as well as estimates of response rates to female students if they had the same preferences as male students, response rates to female students among the professionals preferred by female students, and p-values for a test of whether response rates to female students among the professionals preferred by female students equal response rates to male students among the professionals preferred by male students. Two sets of standard errors are reported: standard errors clustered at the professional level and are reported in parentheses, and bootstrapped standard errors are reported in brackets. The bootstrapped standard errors are based on 50 iterations and cluster at the student-ranker in the step when the model of student preferences is estimated to account for estimation error, and also cluster at the professional-response level when the most-preferred professionals are selected. This procedure yields conservative standard errors.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Effect of Student Gender on Mentions of Work/Life Balance and Workplace Culture, Inclusive of Student Selection

	Work/Life Balance			Workplace Culture		
	(1)	(2)	(3)	(4)	(5)	(6)
Female Student Sent	0.237***	0.233***	0.228***	-0.075	-0.077	-0.074
	(0.075)	(0.081)	(0.081)	(0.063)	(0.072)	(0.069)
	[0.092]	[0.119]	[0.123]	[0.095]	[0.130]	[0.140]
Female Student Preferred		-0.030	-0.046		-0.001	-0.019
		(0.048)	(0.052)		(0.073)	(0.082)
		[0.095]	[0.110]		[0.118]	[0.137]
Female Student Pref x Female Student Sent		0.010	0.026		0.003	0.005
		(0.075)	(0.076)		(0.078)	(0.081)
		[0.155]	[0.154]		[0.164]	[0.162]
Male Pref and Male Sent Mean		0.043			0.170	
Male Pref and Female Sent ($\beta^m - \alpha^m$)		0.275	0.271		0.093	0.096
Female Pref and Female Sent ($\alpha^f + \beta^f$)		0.255	0.251		0.095	0.082
P-value for Experienced F-M diff		0.013	0.011		0.277	0.223
Observations	392	392	392	392	392	392
Message Time/Date	X	X	X	X	X	X
Sender Profile, Race	X	X	X	X	X	X
Student Profile, Race			X			X

Note: This table reports the results of the estimation of the regression specification outlined in equation (6), in which the dependent variable is an indicator for whether a response mentions work/life balance (columns 1-3) or workplace culture (columns 4-6), and the independent variables are an indicator for whether the student who sent the message is female, an indicator for whether the professional was in the preferred set of a female student, and the interaction of these terms, as well as the professional's field, message time/date characteristics, and student profile characteristics. Columns 2, 5 and 8 report results from the baseline specification. Columns 3, 6, and 9 report results from a specification that additionally includes controls for characteristics available on their profile and race/ethnicity of the student selecting the professionals. Columns 1, 4, and 7 report results from the regression specification outlined in equation (1), which relative to the above omits an indicator for whether the professional was in the preferred set of a female student. The bottom panel of the table provides the mean outcome for male students among professionals preferred by male students, as well as estimates of the mean outcome for female students if they had the same preferences as male students, the mean outcome for female students among the professionals preferred by female students, and p-values for a test of whether the mean outcome for female students among the professionals preferred by female students equals the mean outcome for male students among the professionals preferred by male students. Two sets of standard errors are reported: standard errors clustered at the professional level and are reported in parentheses, and bootstrapped standard errors are reported in brackets. The bootstrapped standard errors are based on 50 iterations and cluster at the student-ranker in the step when the model of student preferences is estimated to account for estimation error, and also cluster at the professional-response level when the most-preferred professionals are selected. This procedure yields conservative standard errors.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A Appendix Figures and Tables

Figure A1: Message Templates

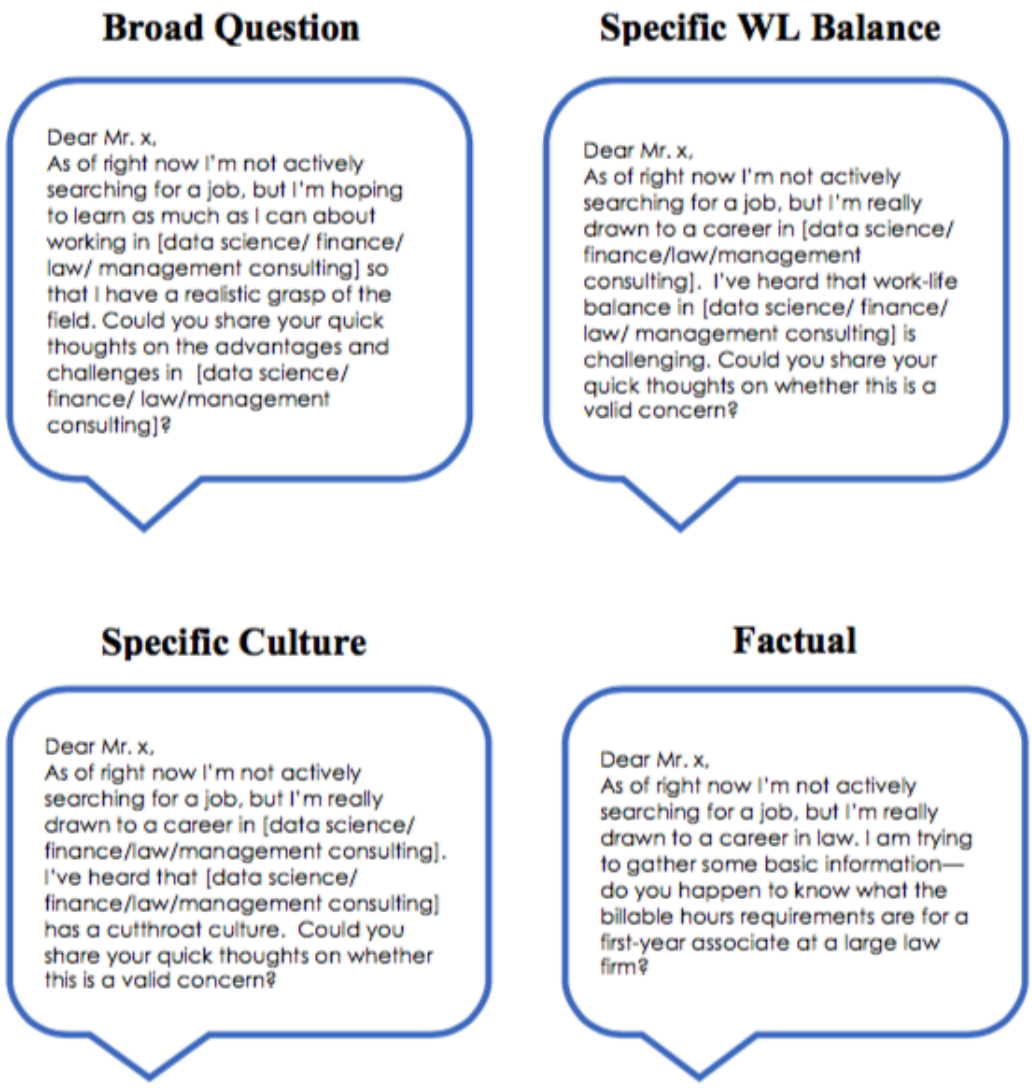

Figure A2: Distribution of Number of Responses

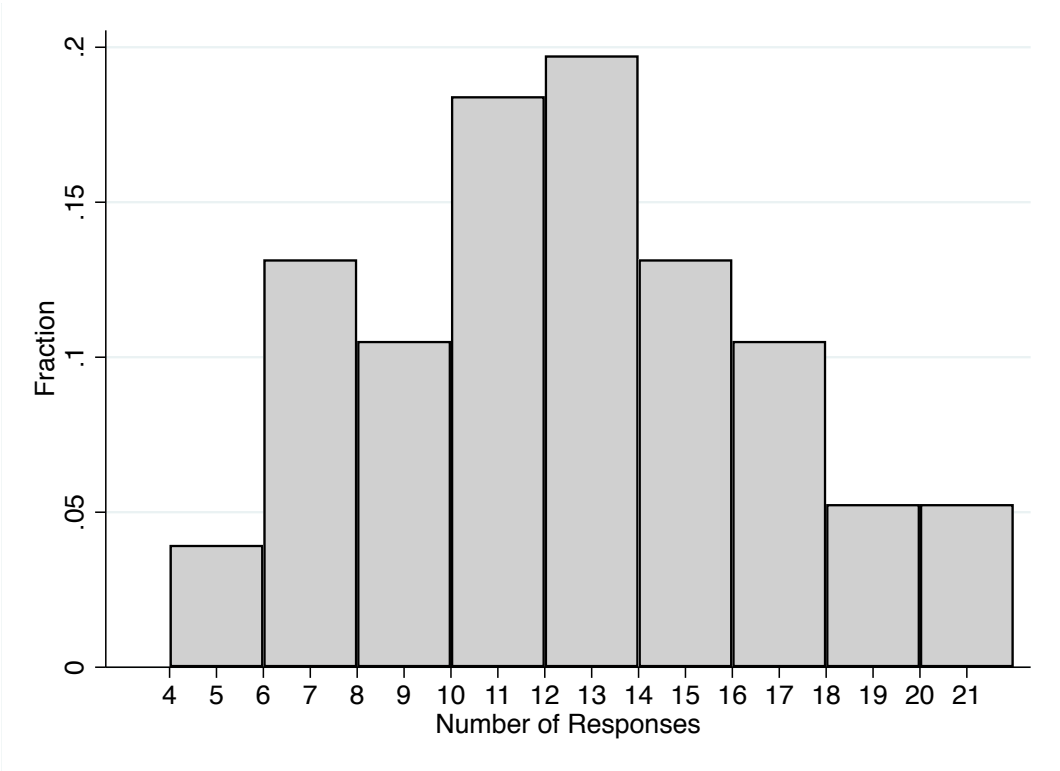

Table A1: Outcome Summary Statistics

	All Messages	Broad	Specific - Work/Life	Specific - Culture	Factual
Response Rate	0.12	0.10	0.14	0.15	0.11
	(0.33)	(0.30)	(0.35)	(0.36)	(0.32)
Response Character Count	434.73	414.39	486.64	429.95	304.18
	(558.77)	(687.34)	(492.37)	(396.61)	(553.03)
Work/Life Balance Mentioned		0.11			
		(0.32)			
Workplace Cuture Mentioned		0.12			
		(0.33)			
Valid concern?					
Yes			0.44	0.16	
			(0.50)	(0.37)	
It depends			0.49	0.54	
			(0.50)	(0.50)	
No			0.07	0.30	
			(0.26)	(0.46)	
Billable Hours Quoted					1989.00
					(77.42)
Observations	7367	3530	1763	1776	298

Note: This table reports summary statistics for the main outcomes, overall and by question type. Means for each outcome are reported, with standard deviations in parentheses.

Table A2: Tests of Randomization

	(1) All Messages	(2) Sent Messages Only
Data Science	0.000	-0.000
	(0.001)	(0.002)
Finance	0.000	-0.003
	(0.001)	(0.004)
Law	-0.000	0.008
	(0.001)	(0.008)
Mgmt Consulting	-0.001	-0.005
	(0.001)	(0.004)
Professional Female	0.004	0.004
	(0.012)	(0.012)
0-249 Connections	0.004	0.004
	(0.007)	(0.008)
250-499 Connections	0.017*	0.021**
	(0.010)	(0.010)
500+ Connections	-0.015	-0.018
	(0.012)	(0.012)
College graduation year	0.311	0.208
	(0.321)	(0.320)
Alumni of Student's College	-0.003	-0.005
	(0.010)	(0.011)
Undergraduation Selectivity Quartile 1	-0.009	-0.012
	(0.008)	(0.008)
Undergraduation Selectivity Quartile 2	0.008	0.008
	(0.009)	(0.009)
Undergraduation Selectivity Quartile 3	0.005	0.008
	(0.011)	(0.012)
Undergraduation Selectivity Quartile 4	0.007	0.006
	(0.008)	(0.008)
Any Graduate Degree	0.008	0.011
	(0.010)	(0.011)
Any Ivy Degree	0.003	0.002
	(0.008)	(0.008)
N	7602	7367

Note: This table reports the results of the estimation of a regression specification, in which the dependent variable is a professional characteristic, listed in the rows, and the independent variable is indicator for whether the student who sent the message to the professional is female. Each entry represents the estimated coefficient from a separate specification. Column 1 reports the results for all messages that were assigned to students. Column 2 reports the results for the subset of messages that students actually sent. Standard errors are clustered at the student level and are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Effect of Student Gender on Mentions of Specific Work/Life Balance Issues

	Duration Typical Workweek		Work Schedule Flexibility		Extent of Travel/Work from Home	
	(1)	(2)	(3)	(4)	(5)	(6)
Student Female	0.054*	0.057*	0.023	0.026	0.015	0.015
	(0.029)	(0.030)	(0.022)	(0.023)	(0.022)	(0.023)
Finance	0.026	0.022	0.021	0.018	0.001	0.002
	(0.025)	(0.024)	(0.013)	(0.015)	(0.018)	(0.018)
Law	0.062	0.057	0.043*	0.037*	-0.004	-0.004
	(0.040)	(0.038)	(0.022)	(0.021)	(0.017)	(0.017)
Mgmt Consulting	0.142***	0.143***	0.089***	0.089***	0.170***	0.171***
	(0.042)	(0.042)	(0.029)	(0.030)	(0.045)	(0.044)
Male Mean	0.047		0.027		0.034	
Observations	363	363	363	363	363	363
Message Time/Date	X	X	X	X	X	X
Student Profile	X	X	X	X	X	X
Student Race/Ethnicity	X	X	X	X	X	X
Additional Student		X		X		X

Note: This table reports the results of the estimation of the regression specification outlined in equation (1), in which the dependent variable is an indicator for whether a response mentions a specific work/life balance issue, and the independent variables are an indicator for whether the student who sent the message is female, the professional's field, message time/date characteristics, and student profile characteristics. Columns 1, 3, and 5 report results from the preferred specification, which also controls for student/race ethnicity. Columns 2, 4, and 6 additionally include controls for student characteristics that may be observable elsewhere online. The omitted field is data science. Standard errors are clustered at the student level and are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Gender Differences in Response Tone (Student Ratings)

	Student more concerned about work/life balance		Student more concerned about workplace culture	
	(1)	(2)	(3)	(4)
	Broad	Work/Life Balance	Broad	Competitive Culture
Student Female	0.034	-0.033	-0.018	-0.004
	(0.027)	(0.048)	(0.015)	(0.034)
Finance	-0.023	0.311***	-0.090***	0.102*
	(0.023)	(0.053)	(0.024)	(0.057)
Law	0.051	0.365***	-0.076***	0.111**
	(0.034)	(0.061)	(0.025)	(0.053)
Mgmt Consulting	0.160***	0.492***	-0.056*	-0.037
	(0.042)	(0.039)	(0.029)	(0.051)
Male Mean	0.114	0.523	0.087	0.293
Observations	3717	2626	3717	2682
Message Time/Date	X	X	X	X
Student Profile	X	X	X	X
Student Race/Ethnicity	X	X	X	X
Additional Student				

Note: This table reports the results of the estimation of the regression specification outlined in equation (1), in which the dependent variable is an indicator for whether a response would make a typical college student more concerned about work/life balance or workplace culture, and the independent variables are an indicator for whether the student who sent the message is female, the professional's field, message time/date characteristics, student profile characteristics, and student race/ethnicity. The question type is listed in each column title. The omitted field is data science. Standard errors are clustered at the student level and are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Effect of Student Gender on Mentions of
Work/Life Balance and Workplace Culture
Accounting for Non-response

	Work/Life Balance		Workplace Culture	
	(1)	(2)	(3)	(4)
Student Female	0.009***	0.008**	0.001	-0.002
	(0.003)	(0.004)	(0.003)	(0.003)
Finance	-0.001	-0.001	-0.023***	-0.023***
	(0.003)	(0.003)	(0.008)	(0.008)
Law	0.002	0.002	-0.028***	-0.028***
	(0.004)	(0.004)	(0.008)	(0.008)
Mgmt Consulting	0.025***	0.025***	-0.014	-0.014
	(0.006)	(0.006)	(0.009)	(0.009)
Male Mean	0.007		0.013	
Observations	3530	3530	3530	3530
Message Time/Date	X	X	X	X
Student Profile	X	X	X	X
Student Race/Ethnicity	X	X	X	X
Additional Student		X		X

Note: This table reports the results of the estimation of the regression specification outlined in equation (1), in which the dependent variable is an indicator for whether a response mentions work/life balance (columns 1-2) or workplace culture (columns 3-4), and the independent variables are an indicator for whether the student who sent the message is female, the professional's field, message time/date characteristics, and student profile characteristics. Messages that do not receive a response are coded as not mentioning these career attributes. Columns 1 and 3 report results from the baseline specification, which includes controls for student race/ethnicity. Columns 2 and 4 additionally include controls for student characteristics that may be observable elsewhere online. The omitted field is data science. Standard errors are clustered at the student level and are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Gender Differences in Responses to "Is work/life balance a concern?"

	Yes		It depends		No	
	(1)	(2)	(3)	(4)	(5)	(6)
Student Female	0.003	-0.008	-0.023	-0.015	0.020	0.023
	(0.071)	(0.074)	(0.076)	(0.078)	(0.034)	(0.035)
Finance	0.216**	0.215**	-0.029	-0.033	-0.186***	-0.182***
	(0.088)	(0.087)	(0.108)	(0.107)	(0.069)	(0.068)
Law	0.315***	0.314***	-0.075	-0.063	-0.240***	-0.251***
	(0.090)	(0.093)	(0.111)	(0.115)	(0.065)	(0.067)
Mgmt Consulting	0.674***	0.672***	-0.427***	-0.422***	-0.247***	-0.251***
	(0.062)	(0.064)	(0.088)	(0.091)	(0.060)	(0.061)
Male Mean	0.427		0.512		0.061	
Observations	211	211	211	211	211	211
Message Time/Date	X	X	X	X	X	X
Student Profile	X	X	X	X	X	X
Student Race/Ethnicity	X	X	X	X	X	X
Additional Student		X		X		X

Note: This table reports the results of the estimation of the regression specification outlined in equation (1), in which the dependent variable is an indicator for whether a response to the specific work/life balance question is one of the categories in the column titles, and the independent variables are an indicator for whether the student who sent the message is female, the professional's field, message time/date characteristics, and student profile characteristics. Columns 1, 3, 5, and 7 report results from the preferred specification, which also controls for student/race ethnicity. Columns 2, 4, 6, and 8 additionally include controls for student characteristics that may be observable elsewhere online. The omitted field is data science. Standard errors are clustered at the student level and are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Gender Differences in Responses to "Is cutthroat culture a concern?"

	Yes		It depends		No	
	(1)	(2)	(3)	(4)	(5)	(6)
Student Female	0.001	0.024	0.053	0.011	-0.054	-0.035
	(0.051)	(0.050)	(0.071)	(0.066)	(0.069)	(0.068)
Finance	0.072	0.082	0.197*	0.205*	-0.269***	-0.287***
	(0.079)	(0.081)	(0.111)	(0.115)	(0.097)	(0.100)
Law	0.136*	0.133*	0.091	0.096	-0.226**	-0.229**
	(0.080)	(0.077)	(0.113)	(0.109)	(0.088)	(0.088)
Mgmt Consulting	-0.035	-0.035	-0.051	-0.037	0.086	0.072
	(0.056)	(0.055)	(0.108)	(0.106)	(0.109)	(0.110)
Male Mean	0.134		0.512		0.354	
Observations	215	215	215	215	215	215
Message Time/Date	X	X	X	X	X	X
Student Profile	X	X	X	X	X	X
Student Race/Ethnicity	X	X	X	X	X	X
Additional Student		X		X		X

Note: This table reports the results of the estimation of the regression specification outlined in equation (1), in which the dependent variable is an indicator for whether a response to the specific competitive culture question is one of the categories in the column titles, and the independent variables are an indicator for whether the student who sent the message is female, the professional's field, message time/date characteristics, and student profile characteristics. Columns 1, 3, 5, and 7 report results from the preferred specification, which also controls for student/race ethnicity. Columns 2, 4, 6, and 8 additionally include controls for student characteristics that may be observable elsewhere online. The omitted field is data science. Standard errors are clustered at the student level and are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A8: Gender Differences in Hours Quoted in Response to Factual Question

	(1)	(2)
Student Female	77.758	149.166***
	(73.706)	(1.931)
Male Mean	1937.500	
Observations	25	25
Message Time/Date	X	X
Student Profile	X	X
Student Race/Ethnicity	X	X
Additional Student		X

Note: This table reports the results of the estimation of the regression specification outlined in equation (1), in which the dependent variable is the hours quoted in responses to the factual question, and the independent variables are an indicator for whether the student who sent the message is female, the professional's field, message time/date characteristics, student profile characteristics, and student race/ethnicity. Columns 1 and 2 analyze only responses that include a numeric value or range. Column 3 and 4 include all responses to the factual question, and impute the minimum billable hours requirement from the NALP website for those responses that do not include a numeric answer. Columns 1 and 3 report results from the preferred specification, which also controls for student/race ethnicity. Columns 2 and 4 additionally include controls for student characteristics that may be observable elsewhere online. The omitted field is data science. Standard errors are clustered at the student level and are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A9: Gender Differences in Response Length

	Broad		Work/Life Balance		Competitive Culture		Factual (Law Only)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Character Count	Log(Count)	Character Count	Log(Count)	Character Count	Log(Count)	Character Count	Log(Count)
Student Female	-10.246	-0.128	10.837	0.094	41.378	0.005	-114.288	-0.323
	(38.899)	(0.108)	(52.098)	(0.157)	(47.904)	(0.112)	(101.655)	(0.589)
Finance	-260.682***	-0.593***	-20.647	0.019	-55.698	-0.170		
	(74.427)	(0.181)	(54.890)	(0.154)	(66.524)	(0.152)		
Law	-210.809***	-0.382**	17.907	0.039	27.385	0.028		
	(78.271)	(0.176)	(57.476)	(0.154)	(69.904)	(0.164)		
Mgmt Consulting	-163.605**	-0.321*	133.915**	0.389***	14.163	0.092		
	(76.373)	(0.177)	(59.993)	(0.143)	(66.047)	(0.141)		
Male Mean	359.547	5.402	414.608	5.691	367.980	5.661	259.889	5.170
Observations	359	359	249	249	262	262	33	33
Message Time/Date	X	X	X	X	X	X	X	X
Student Profile	X	X	X	X	X	X	X	X
Student Race/Ethnicity	X	X	X	X	X	X	X	X

Note: This table reports the results of the estimation of the regression specification outlined in equation (1), in which the dependent variable is the length of the response, and the independent variables are an indicator for whether the student who sent the message is female, the professional's field, message time/date characteristics, student profile characteristics, and student race/ethnicity. Columns 1, 3 5, and 7 analyze the response's character count, while columns 2, 4, 6, and 8 analyze the natural logarithm of the character count. Responses to each question are analyzed separately. The omitted field is data science. Standard errors are clustered at the student level and are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A10: Gender Differences in Other Response Components

	(1) Main Specification	(2) Additional Student Controls
Offers advice of any type	-0.074*	-0.078*
	(0.043)	(0.043)
Explains career paths	-0.038	-0.038
	(0.044)	(0.043)
Mentions analytical aspects of career	0.050	0.051
	(0.042)	(0.041)
Mentions decision-making/responsibility aspects of career	0.013	0.014
	(0.028)	(0.028)
Mentions excitement/impact aspects of career	0.023	0.028
	(0.040)	(0.038)
Mentions interpersonal aspects of career	0.021	0.025
	(0.045)	(0.044)
Compensation	0.052	0.051
	(0.039)	(0.038)
Job Stability	0.033	0.034
	(0.028)	(0.028)
Short v. Long Term Considerations	-0.022	-0.023
	(0.032)	(0.032)
Qualities of Individuals who Like/Succeed	0.043	0.040
	(0.026)	(0.024)
Broadness of Question	-0.025	-0.026
	(0.033)	(0.033)
Info on Job Search	-0.035	-0.037
	(0.029)	(0.029)
Offer to Discuss Further	0.064	0.064
	(0.051)	(0.051)
Decision is Person Specific	-0.024	-0.025
	(0.025)	(0.024)
States Qualifications for Answering	-0.083*	-0.086**
	(0.045)	(0.043)
Education Requirements/Environment	0.031	0.030
	(0.032)	(0.031)
Other attribute	0.011	0.008
	(0.034)	(0.033)
p-value from joint test M=F	0.002	0.001
N	363	363

Note: This table reports the results of the estimation of the regression specification outlined in equation (1), in which the dependent variable is an indicator for whether a response mentions the categories listed in the rows, and the independent variables are an indicator for whether the student who sent the message is female, the professional's field, message time/date characteristics, student profile characteristics, and student race/ethnicity. Column 1 reports results from the preferred specification, while column 2 additionally include controls for student characteristics that may be observable elsewhere online. Standard errors are clustered at the student level and are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A11: Heterogeneity by Professional Attributes

(1) Male	(2) Female	(3) Grad<2010	(4) Grad>=2010	(5) No Ivy	(6) Ivy	(7) Not Alum	(8) Alum	(9) Data Science	(10) Finance	(11) Law	(12) Mgmt Consulting
					Panel A. Response Rate, Broad Question						
0.015	0.006	0.009	-0.002	0.014	0.012	0.018	0.008	0.025	0.015	0.010	0.008
(0.012)	(0.019)	(0.010)	(0.024)	(0.013)	(0.019)	(0.011)	(0.023)	(0.033)	(0.020)	(0.015)	(0.024)
					Panel B. Response Rate, Specific Question - Work/Life Balance						
0.042*	0.046*	0.027	0.089***	0.041**	0.035	0.030*	0.074*	-0.024	0.068*	0.034	0.055*
(0.021)	(0.025)	(0.018)	(0.033)	(0.018)	(0.041)	(0.016)	(0.041)	(0.061)	(0.037)	(0.026)	(0.028)
					Panel C. Response Rate, Specific Question - Competitive Culture						
0.018	-0.016	-0.012	0.011	0.017	-0.033	0.002	0.022	0.006	0.045	0.010	-0.024
(0.023)	(0.031)	(0.023)	(0.039)	(0.014)	(0.057)	(0.018)	(0.038)	(0.061)	(0.031)	(0.032)	(0.040)
					Panel D. Broad Question, Mention of Work/Life Balance						
0.063	0.142**	0.099**	0.042	0.054	0.337**	0.030	0.134***	0.037	0.082*	0.137*	0.064
(0.039)	(0.066)	(0.044)	(0.052)	(0.033)	(0.132)	(0.040)	(0.050)	(0.026)	(0.043)	(0.074)	(0.100)
					Panel E. Broad Question, Mention of Workplace Culture						
-0.022	0.011	-0.030	0.088	-0.038	0.122	-0.072	0.064	-0.065	-0.096	-0.066	0.067
(0.046)	(0.091)	(0.052)	(0.062)	(0.040)	(0.185)	(0.048)	(0.056)	(0.136)	(0.065)	(0.063)	(0.087)

Note: This table reports the results of the estimation of the regression specification outlined in equation (1), in which the dependent variable is listed in each panel title, and the independent variables are an indicator for whether the student who sent the message is female, the professional's field, message time/date characteristics, student profile characteristics, and student race/ethnicity. The column titles list the subsample used for estimation. Each entry in the table reports the estimated coefficient on student female from a separate specification. Standard errors are clustered at the student level and are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A12: Do Female Professionals Treat Female Student Differently?

	Response Rates			Mentions in Broad Question	
	(1)	(2)	(3)	(4)	(5)
	Broad Q	W/L Balance Q	Culture Q	W/L Balance	Culture
Student Female	0.016	0.046**	0.008	0.061	-0.023
	(0.012)	(0.020)	(0.022)	(0.039)	(0.044)
Professional Female	-0.036**	-0.050**	-0.044	-0.007	0.011
	(0.017)	(0.022)	(0.029)	(0.048)	(0.076)
Student Female x Professional Female	-0.006	-0.011	-0.001	0.039	-0.008
	(0.021)	(0.032)	(0.038)	(0.073)	(0.089)
Male Mean	0.101	0.130	0.139	0.067	0.128
Observations	3530	1763	1776	363	363
Message Time/Date	X	X	X	X	X
Student Profile	X	X	X	X	X
Student Race/Ethnicity	X	X	X	X	X

Note: This table reports the results of the estimation of the regression specification, in which the dependent variable is listed in each column title, and the independent variables are an indicator for whether the student who sent the message is female, an indicator for whether the professional who sent the message is female, the interaction of the two indicators, the professional's field, message time/date characteristics, student profile characteristics, and student race/ethnicity. Standard errors are clustered at the student level and are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A13: Effect of Student Gender on Response Rates, By Question Type
Restricting to Students with No Online Presence

	Broad			Work/Life Balance			Competitive Culture			Factual (Law Only)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Student Female	0.031*	0.042**	0.039	0.048*	0.066**	0.086**	0.052	0.043	0.023	0.106***	0.077***	0.066***
	(0.018)	(0.020)	(0.023)	(0.027)	(0.030)	(0.039)	(0.040)	(0.045)	(0.051)	(0.036)	(0.015)	(0.019)
Finance	-0.113***	-0.114***	-0.114***	-0.056	-0.055	-0.053	-0.070	-0.070	-0.070			
	(0.036)	(0.036)	(0.036)	(0.062)	(0.062)	(0.062)	(0.072)	(0.072)	(0.072)			
Law	-0.119***	-0.119***	-0.119***	-0.115*	-0.115*	-0.117*	-0.086	-0.086	-0.081			
	(0.039)	(0.039)	(0.039)	(0.058)	(0.058)	(0.058)	(0.065)	(0.065)	(0.064)			
Mgmt Consulting	-0.074*	-0.076*	-0.076*	-0.058	-0.063	-0.068	-0.054	-0.052	-0.046			
	(0.041)	(0.041)	(0.042)	(0.061)	(0.059)	(0.060)	(0.053)	(0.052)	(0.052)			
Male Mean	0.095			0.115			0.125			0.023		
Observations	1030	1030	1030	511	511	511	520	520	520	87	87	87
Message Time/Date	X	X	X	X	X	X	X	X	X	X	X	X
Student Profile	X	X	X	X	X	X	X	X	X	X	X	X
Student Race/Ethnicity		X	X		X	X		X	X		X	X
Additional Student			X			X			X			X

Note: This table reports the results of the estimation of the regression specification outlined in equation (1), in which the dependent variable is an indicator for whether a message received a response, and the independent variables are an indicator for whether the student who sent the message is female, the professional's field, message time/date characteristics, and student profile characteristics. The sample is restricted to messages sent by students who do not have an online presence. Separate regressions are estimated for each question type: broad, specific - work/life balance, specific - competitive culture, and factual. Columns 1, 4, 7, and 10 report results from the baseline specification. Columns 2, 5, 8, and 11 report results from a specification that additionally includes controls for student race/ethnicity. Columns 3, 6, 9, and 12 additionally include controls for student characteristics that may be observable elsewhere online. The omitted field is data science. Standard errors are clustered at the student level and are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A14: Effect of Student Gender on Mentions of Work/Life Balance and Workplace Culture
Restricting to Students with No Online Presence

	Work/Life Balance		Workplace Culture	
	(1)	(2)	(3)	(4)
Student Female	0.095**	0.136***	0.052	0.060
	(0.036)	(0.030)	(0.043)	(0.046)
Finance	-0.020	-0.026	-0.135	-0.132
	(0.055)	(0.056)	(0.106)	(0.109)
Law	-0.027	-0.040	-0.256*	-0.264*
	(0.070)	(0.077)	(0.129)	(0.140)
Mgmt Consulting	0.068	0.054	-0.112	-0.116
	(0.085)	(0.087)	(0.116)	(0.121)
Male Mean	0.061		0.102	
Observations	110	110	110	110
Message Time/Date	X	X	X	X
Student Profile	X	X	X	X
Student Race/Ethnicity	X	X	X	X
Additional Student		X		X

Note: This table reports the results of the estimation of the regression specification outlined in equation (1), in which the dependent variable is an indicator for whether a response mentions work/life balance (columns 1-2) or workplace culture (columns 3-4), and the independent variables are an indicator for whether the student who sent the message is female, the professional's field, message time/date characteristics, and student profile characteristics. The sample is restricted to messages sent by students who do not have an online presence. Columns 1 and 3 report results from the baseline specification, which includes controls for student race/ethnicity. Columns 2 and 4 additionally include controls for student characteristics that may be observable elsewhere online. The omitted field is data science. Standard errors are clustered at the student level and are reported in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A15: Attributes of Preferred Professionals

	All Preferred Professionals	Female Preferred	Male Preferred
Data Science	0.16	0.09	0.25
	(0.36)	(0.29)	(0.43)
Finance	0.33	0.25	0.44
	(0.47)	(0.43)	(0.50)
Law	0.30	0.41	0.16
	(0.46)	(0.49)	(0.36)
Mgmt Consulting	0.21	0.25	0.16
	(0.41)	(0.43)	(0.36)
Female	0.60	0.88	0.21
	(0.49)	(0.33)	(0.41)
College Graduation Year	2005.16	2004.36	2006.34
	(10.80)	(10.47)	(11.17)
College Selectivity - Admit Rate	0.21	0.22	0.21
	(0.20)	(0.21)	(0.19)
Alumni of Student's College	0.34	0.31	0.39
	(0.47)	(0.46)	(0.49)
Any Graduate Degree	0.75	0.80	0.68
	(0.43)	(0.40)	(0.47)
Any Ivy Degree	0.23	0.27	0.18
	(0.42)	(0.44)	(0.38)
0-249 Connections	0.12	0.09	0.16
	(0.32)	(0.28)	(0.37)
250-499 Connections	0.13	0.11	0.17
	(0.34)	(0.31)	(0.38)
500+ Connections	0.66	0.76	0.53
	(0.47)	(0.43)	(0.50)
Observations	3648	2112	1536

Note: This table reports summary statistics for the preferred sample of professionals, overall and by student gender. Means for each professional characteristic are reported, with standard deviations in parentheses.

B Online Appendix - Details of Manual Classification Categories

Category	Description	Source
Analytical Aspects		
Estimate Quantifiable Characteristics of Products, Events, or Information	Estimating sizes, distances, and quantities; or determining time, costs, resources, or materials needed to perform a work activity	O*NET Work Activity
Get Information	Observing, receiving, and otherwise obtaining information from all relevant sources	O*NET Work Activity
Analyze Data or Information	Identify the underlying principles, reasons, or facts of information by breaking down information or data into separate parts	O*NET Work Activity
Evaluate Information to Determine Compliance with Standards	Using relevant information and individual judgment to determine whether events or processes comply with laws, regulations, or standards.	O*NET Work Activity
Process Information	Compile, code, categorize, calculate, tabulate, audit, or verify information or data	O*NET Work Activity
Interact with Computers	Use computers and computer systems (including hardware and software) to program, write software, set up functions, enter data, or process information	O*NET Work Activity
Interpret Meaning of Information for Others	Translating or explaining what information means and how it can be used	O*NET Work Activity
Decision-making Aspects		
Develop Objectives and Strategies	Establishing long-range objectives and specifying the strategies and actions to achieve them	O*NET Work Activity
Make Decisions or Solve Problems	Analyze information and evaluate results to choose the best solution and solve problems	O*NET Work Activity
Organize, Plan, and Prioritize Work	Developing specific goals and plans to prioritize, organize, and accomplish your work.	O*NET Work Activity
Excitement and Impact Aspects		
Think Creatively	Developing, designing, or creating new applications, ideas, relationships, systems, or products, including artistic contributions	O*NET Work Activity
Update and Use Relevant Knowledge	Keep up-to-date technically and apply new knowledge to your job	O*NET Work Activity
Responsibility for Outcomes	How responsible is the worker for work outcomes and results of other workers	O*NET Work Context
Consequence of Error	How serious would the result usually be if the worker made a mistake that was not readily correctable?	O*NET Work Context
Freedom to Make Decisions	How much decision making freedom, without supervision, does the job offer?	O*NET Work Context

Impact of Decisions on Coworkers/Company Results	What results do your decisions usually have on other people or the image or reputation or financial resources of your employer?	O*NET Work Context
Importance of Being Exact or Accurate	How important is being very exact or highly accurate in performing this job?	O*NET Work Context
Structured v. Unstructured Work	To what extent is this job structured for the worker, rather than allowing the worker to determine tasks, priorities, and goals?	O*NET Work Context
Importance of Repeating Same Tasks	How important is repeating the same physical activities (e.g., key entry) or mental activities (e.g., checking entries in a ledger) over and over, without stopping, to performing this job?	O*NET Work Context
Projects Monotonous/Constantly Changing		Supplemental Category
Interpersonal Aspects		
Communicate with Persons Outside Org.	Communicate with people outside the organization, represent the organization to customers, the public, government, and other external sources. This information can be exchanged in person, in writing, or by telephone or e-mail	O*NET Work Activity
Communicate with Supervisors, Peers, Subordinates	Providing information to supervisors, coworkers, and subordinates by telephone, in written form, e-mail, or in person	O*NET Work Activity
Coordinate Work Activities of Others	Getting members of a group to work together to accomplish tasks	O*NET Work Activity
Developing and Building Teams	Encouraging and building mutual trust, respect, and cooperation among team members	O*NET Work Activity
Establish and Maintain Personal Relationships	Developing constructive and cooperative working relationships with others, and maintaining them over time	O*NET Work Activity
Guide, Direct, and Motivate Subordinates	Providing guidance and direction to subordinates, including setting performance standards and monitoring performance	O*NET Work Activity
Provide Consultation and Advice to Others	Providing guidance and expert advice to management or other groups on technical, systems-, or process-related topics	O*NET Work Activity
Resolve Conflicts and Negotiate with Others	Handling complaints, settling disputes, and resolving grievances and conflicts, or otherwise negotiating with others	O*NET Work Activity
Sell or Influence Others	Convincing others to buy merchandise/goods or to otherwise change their minds or actions	O*NET Work Activity
Coordinate or Lead Others	How important is it to coordinate or lead others in accomplishing work activities in this job?	O*NET Work Context
Deal with External Customers	Job entails work with external customers or the public	O*NET Work Context
Deal with Unpleasant or Angry People	How frequently does the worker have to deal with unpleasant, angry, or discourteous individuals as part of the job requirements?	O*NET Work Context

Face-to-face Discussions	How often do you have to have face-to-face discussions with individuals or teams in this job?	O*NET Work Context
Frequency of Conflict	How often are there conflict situations the employee has to face in this job?	O*NET Work Context
Work with Work Group or Team	How important is it to work with others in a group or team in this job?	O*NET Work Context
Work/Life Balance Aspects		
Duration of Typical Workweek	Number of hours typically worked in one week	O*NET Work Context
Work Schedule Flexibility	Timing of work is flexible/inflexible	Supplemental Category
Extent of Travel/Work from Home	Location of work is flexible/inflexible, including work-related travel	Supplemental Category
Individual Categories that Appear in >5% of Responses		
Explains Paths within Field	Explains various paths within the field	Supplemental Category
Compensation	Mentions pay including salary or bonus	Supplemental Category
Job Stability	Jobs within career path stable/unstable	Supplemental Category
Short v. Long term Considerations	Any time dimension to career path, including whether it positions one well for future jobs or has changing attributes as one gains experience	Supplemental Category
Qualities of Individuals who Like/Succeed	Attributes of people who do well in this career path	Supplemental Category
Broadness of Question	Statement that the question is broad	Supplemental Category
Info on Job Search	Information on how to find a job within the field	Supplemental Category
Implicit/Explicit Offer to Discuss Further	Statement to discuss further (over message, email, phone, etc.) or asks a follow-up question	Supplemental Category
Decision is Person-Specific	Statement that the career decision depends on the person and their attributes/preferences	Supplemental Category
States Qualifications for Answering	Statement of experience in career path with intention of demonstrating that one is/isn't equipped to answer	Supplemental Category
Education Requirements and Environment	Statement of degree requirements and/or description of the attributes of those requirements (e.g. law school is grueling)	Supplemental Category

C Online Appendix - Details of Text Analysis

In this Appendix, we provide details of the Kullback-Leiber Divergence metric and the lexicon-based sentiment analysis.

C.1 Kullback-Leiber Divergence Metric

Throughout this section, we use the term “female corpus" to refer to the set of words (with frequencies) used in all responses to female students. We use the term “male corpus" to refer to the set of words (with frequencies) used in all responses to male students.

When we refer to the distribution of words in a corpus, we refer to the distribution over unique words, where the probability of word j is given by:

$$p_j = \frac{\text{\# of occurrences of j}}{\text{total word-occurrences in corpus}}$$

Note that in this sense, the point estimates do not distinguish between words that occur once in many messages and words that occur many times in a single message: only the total number of occurrences across all messages matter.

Measure of Divergence

In order to compare the differences in language used to respond to male students and female students, we define a measure of divergence, which compares the distribution of words in the female corpus to the distribution of words in the male corpus.

Before defining the measure, we must deal with one critical issue: how to treat words which occur in one corpus but not the other. In our application, the set of words that are not shared across corpi is actually quite large. This can be seen in Table B2. Of the total 3,855 unique words in responses to female students, nearly half are not found in the male responses.

Table B2: Vocabulary Overlap of Responses to Female and Male Students

Analysis	Total Words	Shared Words	Female Only Words	Male Only Words
All	4,817	1,928	1,927	962
Broad	3,045	1,093	1,195	757
Factual	557	135	365	57
Specific Cutthroat	2,444	835	1,123	486
Specific Work-Life	2,402	926	1,020	456

To accommodate this feature of our data, we follow Bohren et al. (2018) and use what we define as the smoothed Kullback-Leiber (K-L) divergence of two corpi. This is the K-L divergence between the two distributions with Lidstone smoothing applied. We use a smoothing parameter of 0.5. The formal definition of our smoothed K-L divergence is given below.

Definition 1 *Given corpus F and corpus M, let V_i denote the vocabulary in corpus i and $C_i()$ denote a function giving the count of a word in corpus i. Then the* ***smoothed K-L divergence*** *of the distributions of F from M is given by:*

$$D_{KLS}(F, M) := \sum_{w \in V_F \cup V_M} p(w) log\left(\frac{p(w)}{q(w)}\right)$$

where:

$$p(w) := \frac{C_F(w) + 0.5}{\sum_{s \in V_F} C_F(s) + 0.5|V_F \cup V_M|}$$

$$q(w) := \frac{C_M(w) + 0.5}{\sum_{s \in V_M} C_M(s) + 0.5|V_F \cup V_M|}$$

We can interpret this measure as the expectation of the logarithmic difference of the distributions, where the expectation uses the female word distribution. In this sense, we are measuring how likely it is that the male observations were taken from the female distribution.

Estimation Procedure

To estimate the K-L divergence metric, we use the definition and replace all probability distributions with their sample analogues. To perform inference we use the bootstrapping procedure outlined in Bohren et al. (2018). This procedure consists of the following:

1. Count the number of responses to male students (N_M) and the number to female students (N_F).
2. For each bootstrap iteration, randomly sample without replacement N_F responses from the full set of responses. Call these responses the placebo female group.
3. Call the remaining N_M responses the placebo male group.
4. Calculate the relevant divergence metric using the placebo groups instead of the true gender.
5. The p-value is the percentage of bootstrap estimates which are less than the point estimate.

To derive what we call p-values clustered at the student level, we perform the following block bootstrap procedure:

1. Count the number of unique male students (N_M) and the number of unique female students (N_F).
2. For each bootstrap iteration, randomly sample without replacement N_F students from the full set of students. Call these students the placebo female group.
3. Call the remaining N_M students the placebo male group.
4. Calculate the relevant divergence metric using the placebo groups instead of the true gender.
5. The p-value is the percentage of bootstrap estimates which are larger than the point estimate.

Consistent with the prior literature, 1,000 bootstrap replications were performed to calculate p-values for each K-L divergence estimate (1,000 replications per p-value).

Data Preparation and Analysis Tools

The sample restrictions are the same as in the main analysis: the 76 students whose names unambiguously convey their gender and who completed the study. The response sample is limited those received within 21 days.

The K-L divergence analyses were conducted using R 3.5.3. The text responses are processed using the packages "stringr" and "quanteda." The command "textstat_frequency" is the main command used to compute word frequencies. Words are defined to be sets of letters separated by spaces. The only processing performed on message text is the removal of punctuation and the removal of the word "x." "x" was used to manually redact messages of identifying information like company and person names. Other than these two processing steps, no other processing was performed. Words are not stemmed and stop words are not removed.

Results

The Kullback-Leiber divergence metric is reported in Table B3. In addition to the overall analysis (denoted "All") the analysis is performed by question type: broad, specific work/life balance, specific competitive culture, and factual.

Table B3 reports point-estimates of the Kullback-Leiber divergence of the male response corpus from the female response corpus. It utilizes the smoothed K-L divergence metric given in Definition 1. P-values are computed using bootstrapping responses. Clustered p-values are computed from bootstrapping students.

Considering all of the responses received, the responses to female students are drawn from a different word distribution than the responses to male students (p-value=0.014). When we look by question type, we observe that the specific questions on workplace competitive culture and work/life balance continue to exhibit gender differences in the word distributions.

Table B3: Smoothed Kullback-Leiber Divergence: Male vs. Female Students

	Responses	K-L Divergence	p-value	Clust. p-value
All	913	1.071	0.066	0.014
Broad	363	0.812	0.358	0.246
Factual	34	3.048	0.563	0.565
Specific Competitive	264	1.248	0.086	0.024
Specific Work-Life	252	1.259	0.080	0.060

C.2 Sentiment Analysis

To measure the tone and emotional content of the messages, we utilize lexicon-based sentiment analysis. Lexicon analyses rely on human-coded databases of words mapped to emotions. The two we utilize are the National Research Council Canada (NRC) lexicon and the Bing lexicon. The NRC lexicon provides 8 emotional categories and two sentiment categories (positive or negative). The Bing lexicon provides two sentiment categories only.[20]. The NRC lexicon contains 6,468 unique words, and each word can have multiple sentiments/emotions (categories are not unique).[21] The Bing lexicon contains 6,785 unique words, and all but three words are uniquely classified as either positive or negative.

For each lexicon and for each sentiment/emotional category in each lexicon, we compute a sentiment score that we call the "sentiment fraction." The sentiment fraction of sentiment j and response i is given by:

$$SF_{i,j} = \frac{\#\text{ words of sentiment } j \text{ in message } i}{\text{total words in message } i}$$

This normalizes sentiment score with respect to message length, and provides a measure of the emotion/sentiment per word in the message. All word counts are counting the number of occurrences of words, not the number of unique words. In this analysis only, we exclude a list of words that are industry related that happen to have sentiment connotation. The full list is given in Table B4. These include words like "lawyer", which in normal conversation would have a negative connotation, but because our experiment involved discussing a career in law, it has a neutral connotation. As a result, these words are excluded from

[20]The website with supporting information is here: `www.cs.uic.edu/~liub/FBS/sentiment-analysis.html`
[21]The website with supporting information is here: `https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm`

both from the sentiment count (numerator) and the count of words (denominator).

Table B4: Removed from Sentiment Analysis

Words
lawyer
attorney
lawsuit
data scientist
analyst
lawyer
counsel
wealth management
director
adjunct
general counsel
law
banker
management consulting
investment banking
data science
finance
scientist
consulting

In terms of vocabulary coverage of the lexicons, out of a total of 4,804 words, 691 words are classified under the Bing lexicon and 936 are classified under the NRC lexicon. We provide a snapshot of the top 10 most frequent words in each sentiment category in Tables B5 and B6.

Table B5: Most Frequently Occurring Words, Bing Sentiments

Negative	Positive
cutthroat	work
challenging	happy
concern	like
hard	good
problems	best
difficult	luck
sorry	great
problem	well
bad	better
stressful	competitive

The sentiment analysis, which includes t-tests (adjusted for multiple hypothesis testing) of the difference

Table B6: Most Frequently Occurring Words, NRC Sentiments

Anger	Anticipation	Disgust	Fear	Joy	Negative	Positive	Sadness	Surprise	Trust
cutthroat	happy	interested	cutthroat	happy	cutthroat	happy	interested	good	happy
challenge	career	bad	challenge	good	challenge	balance	case	hope	good
money	good	delay	case	hope	case	career	problem	luck	hope
bad	time	intense	difficult	luck	demanding	good	bad	money	management
delay	hope	larger	problem	found	problem	culture	late	organization	school
demand	luck	honest	government	pretty	small	job	leave	leave	team
limited	long	criminal	bad	share	dependent	hope	tough	young	found
intense	pretty	finally	change	helpful	litigation	management	delay	expect	pretty
crazy	share	abuse	delay	kind	government	working	limited	guess	legal
honest	enjoy	lie	intense	enjoy	bad	luck	apologize	deal	share

of means, is presented in Table B7. The table compares the mean fraction of words of each sentiment within responses to male and female students. Overall, there are no significant gender differences in the sentiment of responses, nor are there differences in the sentiment of responses to any particular question.

Table B7: Gender Differences in Sentiments of Responses

	All			Broad			Factual			Narrow (Cutthroat)			Narrow (Balance)		
	Diff	p-val	Clust.	Diff	p-val	Clust.	Diff	p-val	Clust.	Diff	p-val	Clust.	Diff	p-val	Clust.
Bing Lexicon															
Negative	0.002	0.505	0.529	0.002	0.631	0.571	0.012	0.554	0.418	-0.003	0.559	0.566	0.004	0.163	0.188
Positive	0.000	0.870	0.889	0.002	0.668	0.702	0.017	0.554	0.418	0.002	0.624	0.579	-0.004	0.455	0.462
NRC Lexicon															
Anger	-0.001	0.950	0.969	0.000	0.990	0.974	0.002	0.688	0.515	-0.003	0.567	0.691	0.000	0.991	0.995
Anticipation	0.001	0.993	0.993	0.005	0.822	0.883	0.023	0.432	0.401	-0.002	0.987	0.989	-0.004	0.810	0.884
Disgust	0.000	0.959	0.970	0.000	0.990	0.974	-0.004	0.236	0.515	0.000	0.987	0.989	-0.001	0.890	0.905
Fear	0.000	0.993	0.993	0.003	0.581	0.433	0.002	0.598	0.449	-0.003	0.728	0.835	-0.002	0.794	0.874
Joy	0.002	0.950	0.949	0.004	0.933	0.903	0.006	0.942	0.949	0.005	0.568	0.633	-0.004	0.808	0.874
Negative	-0.001	0.990	0.993	0.002	0.983	0.961	0.004	0.633	0.515	-0.004	0.490	0.633	0.000	0.991	0.995
Positive	-0.004	0.826	0.752	-0.004	0.965	0.947	0.016	0.834	0.791	0.002	0.987	0.989	-0.012	0.530	0.390
Sadness	0.000	0.993	0.993	0.001	0.990	0.974	0.000	0.995	0.997	0.001	0.987	0.989	-0.002	0.794	0.861
Surprise	0.002	0.746	0.752	0.001	0.990	0.974	0.015	0.688	0.515	0.001	0.987	0.989	0.002	0.890	0.905
Trust	-0.001	0.993	0.993	0.002	0.990	0.974	-0.006	0.942	0.949	0.002	0.978	0.975	-0.006	0.681	0.571

Note:This table displays t-tests comparing the fraction of words from a particular sentiment across female and male students. Negative differences indicate female students received a higher fraction of words from the sentiment than males. p-values are corrected for multiple comparisons (within lexicon). Westfall-Young corrections are performed using the STATA package -wyoung-, with 1,000 bootstrap replications. Column -Clust- reports WY adjusted standard errors clustered at the student level.