NBER WORKING PAPER SERIES

GENDER AND THE DYNAMICS OF ECONOMICS SEMINARS

Pascaline Dupas
Alicia Sasser Modestino
Muriel Niederle
Justin Wolfers
The Seminar Dynamics Collective

Working Paper 28494
http://www.nber.org/papers/w28494

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2021, Revised May 2021

This project reflects an unusual partnership with the Seminar Dynamics Collective, a group of 97 members of the economics community who collected and processed the data used in this study and are considered co-authors, including Corinne Andriola, Victoria Barone, Maryam Blooki, Stephanie Bonds, Nina Buchmann, Drew Burd, Anne Burton, Mrinmoyee Chatterjee, Vittoria Dicandia, Maria Dieci, Karl Dunkle Werner, Holly Dykstra, Luciana Etcheverry, John Fallon, Camille Falézan, Valeria Ferraro, Ellen Fu, Chelsea Garber, Shresth Garg, Anomita Ghosh, Laurie Hakes, Hyoyoung Han, Emma Harrington, Juan Herreño, Kelsi G. Hobbs, Lakshita Jain, Amna Javed, Michelle Jiang, Ariadna Jou, Catherine Michaud Leclerc, Domininkas Mockus, Erica Moszkowski, Philip Mulder, Tuan Nguyen, Urbashee Paul, Dev Patel, Grace Phillips, Xuechao Qian, Rizwanur Rob, Monica Rodriguez, Fernanda Rojas, Arvind Sharma, Rachel Schuh, Rachel Sederberg, Cory Smith, Rizki Nauli Siregar, Melissa Spencer, Anna Stansbury, Ishaana Talesara, Carly Trachtman, Francesca Truffa, Silvia Vannutelli, Joanna Venator, David N. Wasser, Melanie Wallskog, Ashley Wong and Alice Wu. Those not named have chosen to remain anonymous. Special thanks to Jim Poterba for providing coders with access to all of the sessions at the 2019 NBER Summer Institute. Our protocols were reviewed by the Institutional Review Boards at Northeastern University and NBER. The authors would like to acknowledge research funding from Stanford University's Department of Economics. In addition, we are grateful to seminar audiences at Boston University and the American Economic Association's annual conference for comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Gender and the Dynamics of Economics Seminars
Pascaline Dupas, Alicia Sasser Modestino, Muriel Niederle, Justin Wolfers, and The Seminar Dynamics Collective
NBER Working Paper No. 28494
February 2021, Revised May 2021
JEL No. A1,C8,J4,J7

ABSTRACT

This paper reports the results of the first systematic attempt at quantitatively measuring the seminar culture within economics and testing whether it is gender neutral. We collected data on every interaction between presenters and their audience in hundreds of research seminars and job market talks across most leading economics departments, as well as during summer conferences. We find that women presenters are treated differently than their male counterparts. Women are asked more questions during a seminar and the questions asked of women presenters are more likely to be patronizing or hostile. These effects are not due to women presenting in different fields, different seminar series, or different topics, as our analysis controls for the institution, seminar series, and JEL codes associated with each presentation. Moreover, it appears that there are important differences by field and that these differences are not uniformly mitigated by more rigid seminar formats. Our findings add to an emerging literature documenting ways in which women economists are treated differently than men, and suggest yet another potential explanation for their under-representation at senior levels within the economics profession.

Pascaline Dupas
Department of Economics
Stanford University
579 Jane Stanford Way
Stanford, CA 94305-6072
and CEPR
and also NBER
pdupas@stanford.edu

Alicia Sasser Modestino
Northeastern University
310 Renaissance Park
360 Huntington Avenue
Boston, MA 02115
a.modestino@northeastern.edu

Muriel Niederle
Department of Economics
579 Jane Stanford Way
Stanford University
Stanford, CA 94305-6072
and NBER
niederle@stanford.edu

Justin Wolfers
Department of Economics
University of Michigan
611 Tappan St
Lorch Hall #319
Ann Arbor, MI 48104
and NBER
jwolfers@umich.edu

The Seminar Dynamics Collective
pascaline.dupas@gmail.com

1. Introduction

This paper represents the first systematic analysis of the culture of economics seminars. Specifically, we assess the extent to which women economists are treated differently than similarly situated men when presenting their research findings. This question seems particularly pressing, given both the distinctively aggressive culture of economics seminars, and the continuing under-representation of women among the senior ranks of the economics profession.

In winter, spring and summer 2019, we hand-coded data on every interaction between a seminar speaker and their audience in 460 economics talks—including junior faculty recruitment seminars—across most leading economics departments, as well as nearly all talks at the NBER Summer Institute which is a leading annual economics conference. Our rich microdata record the time, duration, type and tenor of each interaction, including the gender and seniority of those making interjections, as well as the gender and many other attributes of the presenter and the research they are presenting.

We document that women presenters are asked 3.8 additional questions ($p<0.01$) relative to men (a 12 percent increase). Accounting for the influence of a range of other factors about the audience, the presenter, the topic, and the coders, reduces the differential to 2.4 questions ($p<0.05$). This disparity appears most pronounced during recruitment ("job market") talks (3.7 extra questions, $p<0.05$) and regular seminars with an external (rather than in-house) speaker (2.6 extra questions, $p<0.05$).

While being asked a different number of questions constitutes differential treatment, it is unclear whether this disparity is helpful or unhelpful, and so we also analyze subjective evaluations of the type and tenor of questions asked. Although we find that women receive a greater number of suggestions and clarifying questions, we also find that they are more likely to be asked questions that are rated as patronizing or hostile.

"Job market talks"—which are part of the recruitment process by which departments decide which economists to hire—are especially high stakes seminars, particularly for those at the start of their careers. As such, the differential treatment of women that we document in this key part of the hiring process is the first evidence potentially linking economics seminar culture with the persistent under-representation of women within the profession.

We supplement our analysis of university seminars with an analysis of talks from the 2019 NBER Summer Institute, which provides insight into how these dynamics play out

in more structured conference presentations. While these presentations are typically less interactive than university seminars, we still detect a gender gap in questions asked, particularly in macroeconomics. Interactions at conferences are typically more highly structured, and our tests of whether differences in the rules (e.g. no questions until the Q&A) can mitigate these gender differences finds surprisingly little evidence that these rules curb the biases we document.

Our data was collected as part of an unusual collaboration with the Seminar Dynamics Collective, a group of (mainly) graduate students who volunteered to analyze seminar dynamics and collect and code relevant data. It was infeasible to recruit "blinded" coders, and so we note that our collaborators are a self-selected group, many of whom likely care more about gender equity than their peers. To allay concern that this may bias our findings, we document remarkably high inter-coder reliability in our raw data, show that our results are robust to controlling for coder fixed effects, illustrate that coders do not have particularly progressive gender views, and document that our key findings are just as evident in the subset of seminars where our data were collected by men as in the subset coded by women.

Before proceeding, it is useful to distinguish between two related research questions. Our analysis focuses on the question of *disparate treatment*, asking whether women are treated differently than men within otherwise-similar seminars. As such, this research can also be read as something of a progress report on whether the economics profession is living up to the ideals laid out in its recently adopted Code of Professional Conduct: to "conduct civil and respective discourse," where "each idea is considered on its own merits," and economists "create a professional environment with equal opportunity and fair treatment for all".[1]

We distinguish this from an alternative—and equally important—question, which is whether seminar culture has a *disparate impact* on women economists, a question taken up by Boustan and Langan (2019). This distinction matters, because even a seminar culture which is gender-blind—that is, a culture that treats men and women equally aggressively—may have a disparate impact, if women find this aggressive or macho culture less welcoming than men do.

The rest of this paper is organized as follows. Section 2 provides some background that motivates our research and surveys the existing literature. Section 3 describes how

[1] The American Economic Association's Code of Professional Conduct was adopted on April 20, 2018. It can be found at this link: https://www.aeaweb.org/about-aea/code-of-conduct.

our data were collected. Section 4 forms the analytic heart of the paper, describing our main findings. Section 5 analyzes a supplementary dataset collected during the NBER Summer Institute in 2019 that complements and enhances our main findings, albeit with less statistical precision in some instances (more in others). Section 6 concludes.

2. Some Background

Our research adds to an emerging literature that has documented that at virtually every margin that has been studied, women economists are treated differently than similarly situated men. Sarsons (2017) finds that women economists receive less credit than their male co-authors when assessed for tenure and promotion. Card et al. (2020) find that journal editors and referees are more likely to reject papers written by women economists than if they were seeking to maximize citations. Koffi (2021) shows that, within subfields, economics papers with female authors are less likely to be appropriately cited by related papers than papers with male authors. Hengel (2018) finds that women experience longer turn-around times and more stringent writing requirements from journal reviewers. Zacchia (2020) shows that some of the ranking methodologies for top economists systematically disadvantage women. And Wu (2020) finds that women economists have been systematically trivialized or even sexualized in online forums. Each of these factors likely contributes to the finding by Chen, Kim, and Liu (2017) that women economists are less likely to be promoted than men, and also to Ginther and Kahn's (2004, 2021) finding that women are less likely to be promoted in economics than in other academic disciplines.

Another motivation for the present study comes from the Code of Professional Conduct adopted by the American Economic Association in 2018. This code targets the profession's seminar culture, noting that the goal of "perfect freedom of economic discussion" imposes "a professional obligation to conduct civil and respectful discourse in all forums." It recognizes the importance of "equal opportunity and fair treatment for all economists, regardless of age, sex, gender identity and expression," and other personal characteristics, while it also articulates a professional responsibility to support "participation and advancement in the economics profession by individuals from all backgrounds, including particularly those that have been historically underrepresented." Yet, according to the recent AEA climate survey (Allgood et al. 2019), the degree to

which one experiences this freedom is far from perfect and varies significantly across different groups. For example, nearly half (47 percent) of female respondents reported that they had not presented their question, idea, or view at their school or place of work to avoid possible harassment, discrimination, or unfair or disrespectful treatment compared to less than one-quarter (24 percent) of male respondents. Similarly, 46 percent of women versus 18 percent of men said they had "not spoken at a conference or during a seminar presentation" to avoid those types of experiences.

While the AEA survey also highlights large differences in the experience of other under-represented groups, our study does not analyze race, ethnicity, sexuality, or disability status because in each case minority groups are so under-represented at top departments that we would lack the statistical power to say much of interest.

These AEA survey results complement other recent studies that have focused more closely on seminars as a potential source of gender disparity in economics. Doleac, Hengel, and Pancotti (2020) have collected data documenting the share of economics seminars given by women or under-represented minorities across 44 leading economics departments. Boustan and Langan (2019) conducted structured interviews with a number of leading economic departments, finding that "departments with better relative outcomes for women are reported to have a less aggressive and more constructive climate in their research seminars."

Studies of other fields outside of economics have also examined how women and men are treated differently when presenting their research, although none appear to be as large or systematic as the present study. Blair-Loy et al. (2017) analyzed videotapes of job talks across five engineering departments at two R1 designated universities, finding that women receive more questions, more follow-ups, and that more of their presentations are consumed by audience speech. Further, they found that the number of questions was related to actions the presenter took which revealed they were rushing to present their slides and complete the talk. Davenport et al. (2014) analyzed presentations at the annual meetings of the American Astronomical Society, finding that women were asked slightly more questions than men were.

Related research has also focused on who is more likely to ask questions in seminars. At the astronomy conference they studied, Davenport et al. (2014) found that women in the audience asked fewer questions than men did. Hinsley, Sutherland, and Johnston (2017) analyzed two international biology conferences, finding that on a per capita basis, men asked nearly twice as many questions as women. Carter et al. (2018) surveyed a

convenience sample of the academic community ranging from undergraduates through graduate students, postdocs, and faculty, about the seminars they attended across a range of fields. Men reported a higher propensity to ask seminar questions than women. While men and women generally reported similar motivations for asking questions, men were twice as likely as women to report being motivated to ask a question because they felt they spotted an error. A majority (58 percent) of their sample—including a majority (60 percent) of women and a near majority of men (48 percent)—reported that they believed that men were more likely to ask seminar questions than women were. In addition, their observation of biology and psychology seminars across a number of countries confirmed that men were more likely to ask questions than women were.

Other studies outside of economics suggest that seminar dynamics are somewhat path-dependent, and Carter et al. (2018) found that women asked proportionately fewer questions when a man asked the first question, or when there were fewer questions overall. Other studies find that the gender of the chair or the overall composition of the audience are important moderators affecting the ratio of questions asked by male versus female audience members (Schmidt et al. 2017).

While there have been few studies on the occurrence of interruptions in conference settings, Miller and Sutherland's (2018) analysis of transcripts from Congressional hearings revealed that women were more likely to be interrupted than men were, and also that women were more likely than men to be interrupted by other women. Jacobi and Schweers (2017) reviewed Oral Arguments from the Supreme Court and found that women were interrupted more often than their male counterparts were, although seniority and political leaning also played a role. The authors note that female justices appeared to learn over time how to behave more like male justices, "avoiding traditionally female linguistic framing in order to reduce the extent to which they are dominated by the men."

3. Data Collection and Summary Statistics: Seminar Sample

Between January 9, 2019 and May 15, 2019, we collected data on interactions between seminar presenters and their audiences in 83 distinct economic seminar series across at 32 universities. Our data represent 460 unique talks featuring 336 presenters (113 women and 223 men). This sample includes 176 job market talks (38 percent) involving 80 job applicants (31 women and 49 men) across 26 universities. Our data collection focused on

leading economics departments, and it includes 20 of the top 30 economics departments.

We developed an online tool to collect detailed information on each seminar, including every interaction between the presenter and the audience. We then recruited a team of collaborators who had the expertise, access and capacity to code the seminars that occurred within their own departments. In the sections that follow, we first discuss the ethical and regulatory issues involved with our data collection as well as the coder recruitment process. Next, we provide a detailed description of the data collection tool and present summary statistics for our sample. Finally, we present evidence about inter-coder reliability and the potential for coder bias.

3.1. Ethical and regulatory issues

The most challenging part of this research came in the data gathering phase. We considered videotaping or audio recording seminars, but quickly learned that in many states this would require opt-in permission from those whose comments would be recorded. Not only did this present feasibility constraints given the scale of data collection involved, but the process of gathering such permissions could have led seminar attendees to alter their behavior. Moreover, it also presented political constraints, as the first few department chairs we approached perceived only downside risk to their individual departments from cooperating, even as they conceded that such research would be valuable for the broader profession. Thus, we were led to collect our data in real time during seminars, coding each interaction as it occurred.

This strategy also presented fewer ethical and regulatory constraints. Research involving human subjects is governed by university Institutional Review Board (IRB) guidelines, and we were careful to obtain permission for all of our research in advance. This process was relatively straightforward, because in general, there is no expectation of privacy among seminar attendees. After all, anyone attending a seminar will observe fellow attendees taking notes, which may include details about what was said and by whom. Our research is simply a more structured form of such data collection. As such, as long as data from each seminar is collected by someone who would normally be invited to attend the seminar and make observations, the data collected are considered "exempt" under IRB guidelines as there is no expectation of privacy and hence no need to obtain informed consent from departments, presenters or seminar attendees.

Even so, we committed to a stronger set of privacy protections. We did not record

the identity of audience members, coding only their gender, seniority, and the timing, tone and type of their comments. We also do not reveal the identities of individual departments or programs when reporting our results. This is because our goal is to assess the state of seminar culture across the economics profession rather than police behavior of any one individual or department. We also chose not to reveal the identities of coders without their explicit permission (even to other coders), to protect their anonymity and reduce the likelihood of retaliation, which was a concern expressed by several of our potential and actual collaborators. Thus, the only personally identifiable information in our dataset is the presenter's name and the title of their talk, both of which were publicly available on the department web page listing the seminar schedule. Even then, we only use this information to create variables that allow us to account for the influence of presenter and paper characteristics (such as home institution of the presenter and JEL code of the paper), and to link to other public information (such as author citation counts, paper publication outcomes, and job placement for the job market sample).

3.2. Coder recruitment

Our data were collected by a group of collaborators we refer to as the Seminar Dynamics Collective to protect their anonymity. Many were recruited through an announcement made at a conference on diversity that drew graduate students from over 30 institutions. Others were recruited by asking for recommendations of potentially interested graduate students from a convenience sampling of faculty at top universities where we did not yet have volunteers. Yet others were recruited through the personal networks of the author team. The result is a convenience sample, albeit one designed with the goal of finding coders at most "top thirty" economics departments.

Our recruitment process resulted in a pool of 77 coders who collected our seminar sample, of whom 73 percent were female, 73 percent were in an applied micro field, and 36 percent were in the 4^{th} year or higher in their Ph.D. program.[2] While it is not unusual for women to be a majority of teams collecting economic data (the U.S. Census Bureau, for instance, was 60% female in September 2020), women are a minority of academic economists and so our coders are not representative of the group we're studying.

The high fraction of women in our team, their interest in diversity-related issues, and the non-blind nature of our study might lead concern that our data collection is biased

[2] We are missing coder characteristics for one coder who did not fill out the registration form.

toward finding disparate treatment by gender. Countering this, Paredes, Paserman, and Pino (2020) find that "exposure to economics causally leads to more gender-biased views," suggesting that perhaps our coders may be biased against finding instances of differential treatment.[3]

Beyond these conflicting conjectures we have collected some useful evidence, and our coders display a pattern of responses on the Harvard Implicit Assumption Test for Gender Career Stereotypes that shows the great majority of them being implicitly biased *against* career women (see Appendix Figure A1). We also conduct several robustness checks at the end of this section to test assess inter-coder reliability and the scope potential bias, finding both high reliability and little evidence of bias. We also control for coder fixed effects in our main analysis and explicitly test (and reject) the hypothesis that female coders systematically assess the gender gaps we observe differently than male coders.

3.3. Data collection instrument and sample characteristics

To collect sufficiently granular data on seminar interactions, we developed an online tool in Qualtrics, a software platform that is commonly used for collecting survey data. The tool presents coders with a series of screens on which they can quickly register relevant observations using a combination of radio buttons with designated choices and comment boxes to add more granular observations. This tool was designed to be used on either a tablet or a laptop so as not to draw attention to the coder during the seminar (thereby reducing the potential for Hawthorne effects). A more detailed description of this tool—including both screenshots and instructions for our coders—is available in Appendix B.

The summary statistics in listed in Tables 1 and 2 represent the first formal quantification of economics seminar culture.

Seminar characteristics

Before the seminar begins, coders use the first page of the Qualtrics tool to record detailed information about the seminar including the time of the seminar, the title of the paper being presented, presenter characteristics (their name, gender, and home institution), and seminar characteristics (duration, whether or not it was a job market talk, number of men and women in the audience, and any "rules" that governed asking

[3] We also note that we had difficulty recruiting coders attending seminar series outside applied micro, especially in macro. If culture is heterogeneous across fields, and those more aggressive or biased fields are less represented in our data, our findings could be underestimates (especially in the regular seminar sample).

questions). We also use this detail to link to other public data describing additional characteristics about the presenter and their paper in order to construct richer control variables.

Table 1 summarizes seminar-level data that we collected about the presenters and the talks both for our main pooled sample, and it also shows detail for the two component sub-samples of regular departmental seminars and job market talks. The final column reports that (perhaps unsurprisingly) there are statistically significant differences between these sub-samples on just about every dimension we measure.

The sample of regular seminars has a lower share of female presenters (32.0 percent) than the job market talks (48.9 percent).[4] Moreover, job market candidates are more likely to be from selective universities (81.3 percent of the job market talks we observe are given by candidates from a top 20 economics department, while presenters from top 20 departments give only 42.3 percent of the regular talks). We will present results for the full (pooled) sample of seminars but also test for differences in the job market and regular samples.

Roughly 53 percent of our regular seminars take place at a top 20 economics department with a higher share (73 percent) among the job market talks since selective institutions tend to have more job openings. Across the regular department seminars, the modal seminar duration is 90 minutes. Roughly three-quarters of the regular talks in our sample are in an applied micro field (a factor partly driven by the interests of our coders), while 17 percent are in macro and 7 percent are in theory or econometrics. Among job market talks, roughly half are in an applied micro field, 24 percent are in theory or econometrics and 26 percent are in macro. On average, 28 people attended each regular seminar (42 for job market talks). Only 7 percent of regular seminars (and 3 percent of job market talks) had any rules in place such as not asking any questions during the first or last 10-15 minutes or only asking clarifying questions initially. Despite most of the seminar series falling under the applied micro category, Figure A1 shows that a wide range of topics were covered across the top-level JEL codes. Women were over-represented

[4] Note that the share of presentations given by women differ from the share of unique female presenters: only 38.8 percent of job candidates observed in our data are female while Table 1 reports they give 48.9 percent of job market talks. The fact that there are fewer women—but that on average each gives more job talks than men—suggests that the apparent robust representation of women in job talks may be driven by an elite group of women being recruited by a wider array of universities. (By contrast, there's no parallel clustering in our sample of regular talks, where women are 32.2 percent of the presenters we observe, and they give 32.0 percent of talks.)

among presentations of papers that were categorized as Financial Economics; Health, Education and Welfare; Development Economics; and Public Economics. Women were under-represented in Macroeconomics; Micro Theory; Industrial Organization; and International Economics (see Appendix Table A1 for statistical tests). Given the large degree of heterogeneity across seminar series within an institution and across topics discussed by presenters, we control for both seminar series fixed and JEL code group fixed effects in our main analysis.

Interaction-level data

Our coders used our online survey tool to record data—in real-time—about every interaction between the presenter and their audience during each seminar. This includes objective measures such as the start and end time of each interaction, the number of interactions, who asked the question (e.g., male or female, professor or student), and whether the question was answered, deferred, ignored, or interrupted. We also collected more subjective evaluations of each interaction coding both the type of question (e.g., comment, criticism, suggestion, clarification, or follow-up) and the tone of the question (e.g., supportive, patronizing, disruptive, demeaning, or hostile).

Table 2 reports summary statistics, again showing both the pooled sample and the two component sub-samples, noting significant differences between the two. On average, roughly 26 questions are asked during a regular economics seminar and 35 questions are asked during a job market talk. For a 90-minute seminar, this represents one interruption every 3.5 and 2.5 minutes, respectively—although interruptions are not uniformly distributed during the time allotted. Moreover, there is considerable heterogeneity with the number of questions ranging from a low of 5 to a high of 69 for any given seminar. There are 3.6 times as many questions from men as from women during regular seminars—and 7.6 times during job market talks—despite men only outnumbering women roughly 2 to 1 in attendance (and 3 to 1 in job talks).

Very few questions are deferred or ignored, suggesting that questions can potentially take up a lot of time during a seminar. On average, in both types of talks, it only takes about 7 minutes before the first question is asked with roughly two questions asked within the first 10 minutes. Overall, questions and responses take up about 24 minutes of regular seminars and 31 minutes of job market talks. Again, for a typical 90-minute seminar, this represents about one-third of the time being taken up by questions. While more of this time is taken up by speaker answers during the regular seminars, it is more evenly split

between the questioner and the presenter during the job market talks. In terms of the type of questions, roughly 35 percent of all questions in regular seminars (37 percent in job market talks) are classified as clarifications, followed by another 17 percent (13 percent) that are classified as comments. Suggestions, follow-ups, and criticisms each account for 10 percent or less in both regular seminars and job market talks—perhaps countering the reputation that economics is as an overly critical discipline.

What about the tone of the question? Here, we specifically asked coders not to code the intention of the person asking the question, nor how it was taken by the speaker, but rather, the coder's assessment of the tenor of the interruption in a scientific setting. Coders had the option to leave this assessment blank and were instructed to code the interaction only if they thought it was warranted. They could also code an interaction as having more than one tone. For example, an interaction could be supportive, patronizing, or both. As shown in the coder guide presented in Appendix B, we defined these terms for coders as follows:

- Supportive: For example, I provide the speaker with a great example they can use. Or provide an answer to a problem. Or I tell them why I find their insight useful.
- Patronizing: A question or comment that may be apparently kind or helpful, but betrays a feeling or sense of superiority over the speaker. A question or comment could be both supportive and patronizing if the interjection acts as if the speaker can't answer themselves.
- Disruptive: Here we think of interactions that disrupt the flow of the seminar, maybe shifting the talk into a completely different direction, away from the speaker and their research.
- Demeaning: A question or comment that – in some measure – causes the speaker to lose their dignity or the respect of others. A demeaning question or comment is less about the scientific point being made, and more about shifting the focus to the speaker and undermining their status as an expert.
- Hostile: A question or comment that is unnecessarily antagonistic, aggressive, confrontational or combative. Hostility describes an aggressive interaction, one that you may not want to encounter as a speaker. Hostility is not required to make a scientific point.

Most interactions were not given a subjective assessment by the coder: only 4 questions out of 26 on average for regular talks, and 3 out of 35 for job market talks. Among rated interactions in regular seminars, 87 percent (45 percent for job market talks) were coded

as supportive, 8 percent (35 percent) as patronizing, 6 percent (16 percent) as disruptive, 1 percent (7 percent) as hostile and 1 percent (5 percent) as demeaning.

Final observations

At the conclusion of the seminar, coders were asked to report some final seminar-level observations. This included using a Likert scale to assess both the degree to which the overall tone of the questions asked were unfair and whether the presenter seemed confident. Coders also assessed the degree to which attendance was lower or higher than usual for that seminar series as well as whether there were any particularly disruptive audience members and their gender. There was also an open-ended comment box for coders to note any further comments or impressions. Table 2 indicates that for one out of every 10 job talks, coders thought the questions were unfair overall. Even more striking is that roughly one in 5 job talks had a particularly disruptive audience member, one in 10 job talks had more than one disrupter in the audience, and the disrupters were mostly male.

3.4. Coder Reliability

We perform several robustness checks to assess the reliability of our data. Panel A of Figure 2 leverages the fact that in 84 seminars two coders were present, and so it shows how each coded the number of questions asked during the seminar. While there are some outliers, the degree of correlation between the two coders is quite high (correlation=0.92).

Panel B of Figure 2 presents a more detailed case study, showing the minute-by-minute seminar coding timeline for two seminars (one with a woman presenting, the other with a man), where data was recorded by both a male and female coder from the same institution. The similarity of the coding of the timing of the interjections, the questions, the answers, and the back and forth across the two codes is striking, suggesting that coder discretion does not play much of a role in our data.

3.5. Job market candidates: mid-term outcomes

We complement the dataset with mid-term outcomes data for the job market candidates. In December 2020, close to two years after the job market season considered, we checked the webpages of all job market candidates in our dataset and coded the candidates' placements, the rankings of their institutions, whether they had received a

post-doctoral fellowship prior to starting a tenure-track position, whether their job market paper was already published or at the "revise and resubmit" stage in a top journal, etc. We present those outcomes in Table A2.

Female job market candidates appear to have had higher quality papers on average, if publication success is any indication (see bottom row), though the difference is not statistically significant given our limited sample size. Of course, job placement and, to some extent, paper placement, could be directly impacted by interactions during job market talks, so these measures are potentially endogenous. Even so, we show in a later appendix (Table A8) that controlling for these measures of candidate and paper quality does not alter the results qualitatively—if anything, it *increases* the gender differentials we observe.

3.6. Regular seminar speakers: citations

We complement the dataset with citation data for the regular seminar speakers. In April 2021, we looked up the google scholar profiles of all speakers and, when available, collected their citation counts as of December 2018 and their citation counts in 2018 alone. 89% of presenters in regular seminars had a google scholar profile. Table A3 reports summary statistics for the citation data by gender. Male speakers are more likely than female speakers are to have a google scholar profile (91 percent compared to 84 percent) and, on average, they had higher citation counts at the end of 2018 (5,685 compared to 4,120 citations), although that difference is not significant. We show that controlling for these measures of speaker quality for the regular seminar talks sample does not affect our main results.

4. Results from Departmental Seminars and Job Market Talks

We seek to explore the degree to which female presenters experience disparate treatment during seminars relative to males using the observational data collected on each interaction between the presenter and the audience members. Given that little is known about seminar dynamics in economics, we analyze both objective (quantitative) as well as subjective (qualitative) indicators. In terms of quantitative outcomes, we examine the number of questions asked, how questions were handled by the presenter (e.g., answered, deferred, or ignored) and the total amount of time spent on questions during the seminar.

In terms of qualitative outcomes, we assess the type and tone of the questions asked as well as the degree to which audience members were disruptive.

Given the considerable heterogeneity across presenters, seminar settings, and coders, we assess whether female economists experience disparate treatment during economics seminars using the following rich OLS specification:

$$Y_{p,s,c} = \alpha + \beta_1 Female\ Presenter_{p,s} + D_s'\beta_2 + \beta_3\delta_s + \beta_4\gamma_c + \beta_5\eta_p + JEL_{p,s} + \varepsilon_{p,s,c}$$

Where for presenter p in seminar series s where data are collected by coder c:

$Y_{p,s,c}$ is the outcome of interest (e.g., number of questions asked);

$Female\ Presenter_{p,s}$ is an indicator variable for whether the presenter is female;

D_s is a vector of talk level controls comprising indicator variables for official seminar duration in minutes (e.g., 60, 75, 80, 90 minutes), whether the seminar is internal (that is, the presenter is from institution hosting the seminar) and whether the seminar is a graduate student workshop (based on the name of the seminar series)

δ_s are seminar series fixed effects (a combination of seminar series title and institution, so there's a single indicator for say, the Michigan labor seminar);

γ_c are coder fixed effects;

η_p are indicator variables describing the ranking of the presenter's home institution;

JEL_p are indicator variables for each of the top-level JEL codes describing the field of the presenter's paper; and

$\varepsilon_{p,s,c}$ is a stochastic error term

The coefficient of interest in equation (1) is β_1, which measures the differential between female and male presenters, controlling for the duration of the talk as well as seminar series, coder, home institution group and JEL fixed effects. We have too many singletons in terms of presenter home institutions to include institution-specific fixed effects for all, so we instead create group home institution fixed effects by categorizing institutions by type and rank, yielding 5 "home institution" groups that we include as fixed effects (the list is shown in Table A4). We also group JEL code categories with few observations, resulting in 7 "JEL code" groups that we include as fixed effects (the list is shown in Table A5).

If β_1 is positive and significant for an outcome such as the number of questions asked, this would indicate that women receive more questions than men do during talks. Whether or not this is harmful to women remains an open question, but it would certainly

be an indicator of *disparate treatment* of women presenters.

Note that our unit of observation is a talk-coder pair—a specific talk coded by an individual coder. In some cases, a talk may have been recorded by more than one coder. As such, we weighted each observation by the inverse of number of coders recording a given talk so that each talk is given equal weight.

For completeness, we report robust standard errors clustered in three distinct ways: the presenter level (since we sometimes observe the same presenter presenting in different seminar series—often with the same paper), the seminar series level, and the talk level (since we have some talks recorded twice by two different coders). Each approach yields rather similar estimates of the relevant standard error, suggesting that there is not much correlation of the error term across observations (at least along each of these dimensions).

We present results for the full sample and also separately for the sample of job market talks and the sample of regular seminar talks. We use OLS primarily, but as needed, we use an alternative nonlinear method to relax the linear functional form assumption.

4.1. Objective (quantitative) indicators

Number of questions asked

We start by analyzing the most straightforward outcome, which is the number of questions asked, on average, to women and men when presenting their research. The first column of Table 3, Panel A shows that on average, women receive roughly 3.8 more questions during a seminar compared to men and the effect is significant no matter how we cluster the standard errors. This coefficient is virtually unchanged when adding controls for seminar duration in column 2. The gender gap remains significant but decreases to 2.4 when controlling for seminar series fixed effects in column 3, a decline on the order of one standard error. Subsequent specifications add fixed effects describing the presenter's prestige (their home institution), the field of their paper, and the identity of the coder, and the coefficient of interest remains remarkably stable as we add this rich set of controls.

Panel B of Table 3 breaks down our results by talk types.[5] The gender differential in

[5] We do not include coder fixed effects when we split the sample between regular and job market talks, because we lose a significant number of singleton observations when we do so (i.e., when a coder coded multiple regular talks but just one job market talk, or vice versa).

the number of questions asked is greatest in the sub-sample of job market talks, at 3.8 with the battery of fixed effects (column 2).[6] This suggests that the differences we document may impact how the profession assesses candidates for hire. The gap in number of questions in external regular talks (when the speaker comes from another institution) is 2.6 (column 4). We have a small sample of internal talks where speakers present at their home institution—these are mostly student workshops—and in these cases we see a reversal in the gender gap: women are asked fewer questions on average, though the confidence intervals are very large and we cannot reject the null of no difference. For the rest of the analysis, we pool external and internal talks into one "regular talks" group since we have too few internal talks to analyze them separately.

We perform several robustness checks to ensure that our results are not driven by outliers. Figure 3 shows that the distribution of the number of questions asked during seminars presented by women is clearly shifted to the right relative to the distribution for men. Outliers appear not to play much of a role for either gender, although this mean shift combined with the greater variance in the number of questions asked of men means that women are noticeably less likely to receive few questions. We have also re-estimated our main findings using a Poisson Quasi-Maximum Likelihood Estimator (QMLE) instead of OLS, and Table A6 Panel A shows that this yields nearly indistinguishable results.

We also assess whether the demographic composition of our coders plays a role, and Table A7 reports results that includes interaction of our coefficient of interest with the coder's gender. We find that the measured differential in the number of questions asked is quite similar whether our data come from the observations of male or female coders. Subsequent results in Table A7 shows that our findings remain remarkably constant across other coder characteristics—such as whether our coders are early or late in their doctoral studies, or their primary field of study.

Could the gap in the number of questions asked be driven by differences in the audience? After all, talks by women attract a larger attendance on average (Table A1). In Table A6 Panel B, we add controls for total attendance. The magnitude of the coefficient on "female" decreases by less than 0.5 questions, suggesting that while the greater attendance may be one of the mechanisms (an extensive margin effect), most of the gender gap in questions asked seems to come from an effect on the intensive margin.

Could the gap be driven by differences in paper or presenter quality? This question is

[6] There are 4 non-rookie candidates in the job market sample. Dropping these leaves the results unchanged.

difficult to answer absent objective measures of quality. One response is to note that our job market sample is already highly selected in a way that minimizes variability, because all job market candidates must be above the "bar" to be possibly employable by the hiring institution (and for mid-tier departments, they must not be above some bar that would suggest they're unattainable). Another response is to controls for presenter and paper quality. Table A8 reports results for our sample of job market talks which add controls for a rough proxy of the paper's quality (whether subsequently on track to be published in a top five journal), as well as the presenter's quality (whether they earned a "top ten" tenure track job). If anything, these controls somewhat amplify the gender differential in questions asked. We can perform a similar test for our sample of regular seminars by controlling for the presenter's number of google scholar citations as of December 2018. As Table A9 shows, adding citation controls in a variety of ways does not much change estimates of our coefficient of interest.

Does disparate treatment of women during seminars vary by field within the discipline? After all, there is not one single "economics culture," but often different norms within different fields. For example, Figure 4 shows that fewer questions are asked in econometrics and theory seminars, a fact that our main specification controls for, because we control for fixed effects for each seminar series at each institution. But here, our question is whether these different cultures shape our observed gender differentials. As such, in Table A10, we interact our coefficient of interest with coarse indicators of the type of each seminar (grouping them into applied micro vs. macro vs. other vs. job market seminars—the latter being typically attended by all fields). These results are relatively imprecise—the effects are large in most fields but not always statistically significant, and differences between fields are hard to detect because we have a relatively small sample of seminars that are neither applied micro nor a job market talk (N=75, see the breakdown in Table A11).[7] As such, we defer further exploration of differences across fields until we explore the NBER sample in section 5, which has a much greater representation of macro and finance, in particular.

[7] Indeed, we only have 7 macro seminar series in the sample, one of which has no female speakers; only two have more than one female speaker (two and four). This means that the specification in columns 3-6 are estimated off of extremely few observations (two seminar series and 9 female talks). The gender gap is very large for macro absent any fixed effects, but as we limit ourselves to the variation that comes only within seminar series, the sign flips. We face some of the same limitations when we attempt to estimate heterogeneity by JEL code in Table A12.

Other objective indicators

Who is asking additional questions of female presenters during seminars? In Table 4, we perform the same analysis as before, but now the dependent variable is the number of questions asked by people of a specific gender × seniority group. As such, this allows something of a decomposition of who is asking these extra questions of women. We find that the extra questions appear to come from male and female faculty in roughly equal measure (although the coefficient on the former is not statistically significant, reflecting the much larger standard error).

The gender of the audience members asking more questions is however very different across types of talks. In regular seminars (column 4), the additional questions are almost entirely due to extra questions asked by women in the audience (column 4), which is all the more striking given that they only make up a third of the audience (Table 1). By contrast, in job market talks, 75% of the extra questions asked of women come from men in the audience, which roughly matches their audience share.

One interesting finding is that female presenters receive more questions from female students, and this effect while quantitatively small (graduate students ask very few questions), is statistically significant, and might suggest that female presenters play an "empowerment" role for young female scholars.

Just because someone asks you a question does not mean that you answer it. If women receive more questions than men do, do they in fact answer more questions? Rows (6) through (8) in Table 4 show that the answer is clearly yes as the number of extra questions deferred or ignored by female presenters is not significantly different from 0.

If women answer more questions during seminars than men, does this add up them spending a greater share of their presentations on questions? This could be important if it leaves women with less time to highlight key findings, or it leads them appearing rushed towards the end of their seminars, or otherwise leads their presentations to be less convincing. Table A13 turns to analyzing time spent on questions, rather than just the number of questions asked, and we find—perhaps surprisingly—that women do not spend significantly more of their seminar time in questions. One possibility is that audience members are less likely to consolidate their comments/questions to female presenters than they are to male presenters, generating more interruptions. These interruptions do not seem to come disproportionately early as we also find that women are not more likely to receive their first question earlier than men are and do not receive more questions in the first 10 minutes of the seminar compared to men (see Table A13).

4.2. Subjective indicators

Economics is known for its aggressive seminar culture and most economists can probably recall a particular seminar where they felt uncomfortable as either a presenter or an audience member. In this section, we measure the degree to which female presenters face harsher criticism and/or greater hostility from audience members compared to male presenters.

In terms of the type of questions, recall from Table 2 that in the sample of regular seminar talks roughly 35 percent of all questions are classified as clarifications, followed by another 17 percent that are classified as comments. Suggestions, follow-ups, and criticisms each account for just under 10 percent each. The bottom panel of Table 4 explores the degree to which female presenters are more or less likely to receive each of these question types. (And Table A14 further breaks this down by the gender of the audience members asking questions.) There is no evidence that women receive more criticism than men do. With the full battery of controls and fixed effects, female presenters are more likely to receive both suggestions and clarifications—about 0.6 additional suggestion and 1.3 additional clarifying questions.

What about the tone of the question? Table 5 assesses the degree to which women are more likely to receive questions with supportive, patronizing, disruptive, demeaning, or hostile tones. We find some evidence that women are more likely to receive patronizing questions. Women are also more likely to receive questions coded as hostile—even when controlling for seminar series, coder, home institution group, and JEL code group fixed effects—and those additional hostile questions seem to be coming almost entirely from male audience members. Aggregating across negative tones (questions which are patronizing, disruptive, demeaning or hostile), women receive 0.5 more such questions, three quarters of which seem to be coming from men. Comments written by coders (shown in their entirety in Appendix C) support this finding. One coder noted of a job market talk,

> "*Despite warning the room that she was running out of time, the questions continued. Nearing the end, one male professor insisted on an answer to a previous question with which he was unsatisfied, continued to speak over her for a time when she tried to move on, and instigated an entire corner of the room to talk over her. There was no time left at the end for Q&A, and despite cheery responses and confidence throughout interruptions, this closing "question" (disruption) seemed especially demoralizing.*"

At the end of the talk, our coders rated the overall fairness of the questions asked using a Likert scale. Figure A3 shows that female presenters are less likely to be asked questions that were rated either "extremely fair" or "somewhat fair", while they were more likely to be asked questions that were assessed as being "somewhat unfair". (The category of "extremely unfair" was only used once.) These differences are particularly stark during job market talks.

5. Evidence from the NBER Summer Institute

In the summer that followed our main data collection covering university seminars, we collected a supplementary dataset from the National Bureau of Economics (NBER) Summer Institute.

Background on the Summer Institute

The NBER is an important facilitator of economic debate, and arguably, the leading convener of top economists around the world. Each year it holds its Summer Institute, which is a month-long series of invitation-only conferences, each run as a series of "program meetings. (e.g. "Monetary economics").

We obtained approval from the NBER for coders to sit at the back of each program session to collect data. The tool we used was modified in order to accommodate NBER's requests that the data be even more comprehensively anonymized. In particular we were not permitted to record the identity of the specific program meeting we were observing. And so in contrast to our earlier analysis of departmental seminars in which we controlled for each seminar series at each institution—effectively holding the field, audience, room and seminar format and group norms constant—we were only able to account for whether each meeting was in one of three broad areas: micro, macro or finance.[8] In addition we were not permitted to record the names of individual presenters, nor the titles of their papers, which means that we cannot link to other data describing characteristics of the presenter or their paper. We also agreed not to record potentially provocative information on the negative "sentiment" of the question or comment (patronizing, hostile, etc.). Instead, we were permitted to code whether the interaction generated by a question or

[8] Our coding into these broad categories follows the scheme compiled by Chari and Goldsmith-Pinkham, 2017, as agreed upon with NBER.

comment was particularly collegial, constructive, or valuable.

Despite these constraints, the NBER sample has important strengths. In particular, the Summer Institute includes workshops across a great number of different fields in economics, which gives us the opportunity to explore differences across fields. And the Summer Institute tends to involve more structured discussion, which gives us both an opportunity to explore whether the differences we observed in more free-flowing seminars persist in more structured conference presentations, and also to explore the mediating influence of the different rules that different groups use to govern how and when questions may be asked.

Presentations and audience interactions at the NBER Summer Institute differ from those in the university department seminars that we analyzed, in several marked ways. While most university seminars are 90 minutes (or close to) in length, the presentation time at the Summer Institute varies dramatically across programs, with a mean of just 39 minutes (see Table 6). A number of programs include not only formal presentations but also formal discussants. In such cases, the presentation is shorter, and the audience tends to limit questions ahead of the discussant's remarks. Some programs make such restraints an official rule, postponing all but clarifying questions until a formal Q&A discussion period at the end. Other programs impose a moratorium on questions in the first 10 or 15 minutes (or up until the end of the introduction), and/or in the last 5 or 10 minutes. Such rules are more common than not: of the 447 talks coded in our data, only 36 percent have the sort of laissez faire approach favored by regular university department seminars (no discussant, no Q&A, and no moratorium on questions). Instead, 44 percent of NBER seminars have a discussant and Q&A session at the end. Another 7 percent have a Q&A without discussant and 5 percent have a discussant without Q&A. Only 10 percent have a moratorium on questions at the beginning but no Q&A. All told, these data are better thought of as representing conference presentations than regular seminars, and they lack much of the give and take of a seminar discussion.

This sample is also different because all NBER Summer Institute participants were informed that our study was taking place. Indeed, the week before the Summer Institute started, NBER President and CEO Jim Poterba emailed registered participants to raise "three conference-related issues." The first issue concerned Twitter, and the need for audience members to seek consent from presenters before distributing photos of slides. The two other issues were related to our project, one directly and one indirectly. Specifically, Poterba wrote:

> "*A team studying seminar culture will be collecting data on presenter - audience interactions during Summer Institute sessions. The data collectors will be recording data on the timing and nature of questions and other interactions in an anonymized format. Prospective summary statistics will aggregate the findings from multiple meetings. This study has been reviewed by the NBER's Institutional Review Board, and I look forward to learning in a systematic fashion about this aspect of our conferences. Finally, please be mindful of the NBER's Conference Code of Conduct,*[9] *which is designed to foster the lively and productive exchange of scientific ideas in an environment that is free of harassment and discrimination."*

In addition, program directors were free to make an announcement at the beginning of their program's conference to remind the audience of our data collection efforts, and coders noted that this occurred in at least two programs.

Despite logistical constraints (we obtained approval only a week before the conference began), we were able to code presentations at 48 of the 51 program meetings, yielding a total of 443 talks, of which 122 (28 percent) were presented by women. This supplementary dataset was collected by a group of 29 coders, of which only 4 had previously participated in coding our university seminar sample. These coders were recruited from local Boston/Cambridge institutions, and 52 percent are female, 83 percent specialized in an applied micro field and 31 percent were in 4th year or higher in their Ph.D program (see Appendix Table A15).

Summary statistics of the NBER sample

Table 6 presents summary statistics for the talks in our NBER sample. The first column shows the mean of each variable for the whole sample. Only 28% of these NBER presentations were given by women (compared with 32% of the regular seminar sub-sample, and 49% in the job market talk sub-sample). NBER talks also have a larger audience, and on average have 63 people in attendance of whom 28% are women. The average number of questions per session is only 14, commensurate with the shorter duration and format constraints discussed earlier.

One distinct advantage of the NBER sample is the more balanced representation

[9] https://nber.org/COI/ConferenceCodeofConduct.pdf. As far as we know, the NBER Conference Code of Conduct was written in anticipation of the 2019 Summer Institute and had not been circulated prior to any other NBER conference. Building on this, at its September 2019 meeting, the NBER Board of Directors adopted a Code of Professional Conduct for NBER-affiliated researchers: https://nber.org/COI/CodeofConductPolicy0919.pdf

across the major fields, and so Table 6 is also structured to highlight differences across fields. The first column shows the mean of each variable (the row title) for the entire NBER sample, while the next column shows the mean for the macroeconomic sub-sample, and the final two columns show how the averages in micro and finance talks deviate from the macroeconomic base case. Reading across the first line, for instance, reveals that women give only 18.7% of all macroeconomic presentations, but their share in micro is 16.3 percentage points higher (and their representation in finance is 7.1 percentage points higher than in macro). This echoes the pattern described in Chari and Goldsmith-Pinkham, 2017. The next set of analyses reveal that Finance talks have a very different format than regular seminars, with all Finance talks having a discussant and/or a Q&A session. Interestingly, both Micro and Finance talks involve significantly fewer questions than Macro and the share of questions that are neither valuable, nor constructive, nor collegial is much lower for Micro talks.

Analyzing gender differences in the NBER sample

Tables 7 and 8 present our analysis of whether there are differences in how male and female speakers are treated at the NBER Summer Institute, using a regression format analogous to that used in the earlier analysis. The one exception is that we cannot control for program fixed effects to control for the specific audience, format and norms of particular programs as we did when using seminar series fixed effects in the spring seminars sample. Instead, we use a cruder set of controls for "Field × Format" fixed effects, which we construct as a saturated set of interactions of our three fields (Micro, Macro and Finance), whether the session was joint between multiple programs, and format variables based on whether there was a discussant, a Q&A session at the end, or a moratorium on questions at the beginning. This yields 15 categories associated with a specific field and format and with at least one female and one male presenter (and each category has a mean of 26 observations).

Each row in Table 7 corresponds to an outcome of interest, the first being our primary outcome from before: the total number of questions asked of the presenter. Column (1) shows the overall mean and standard error, column 2 shows the coefficient on "Female Presenter" for the full sample, and columns (3) through (8) show results for separate regressions in a variety of subsamples, by field and presenter characteristics. For example, the estimate in row (1), column (2) of Table 7 shows that female presenters were asked an additional 1.3 questions on average compared to male presenters (p-value=0.156).

While this is not statistically significant, it is a differential of +9.3% (compared to the mean of 14 questions), which is comparable to the +12 percent rise observed in the spring seminars sample (where we observed a differential of 3.8 questions compared to a mean of 30). Analyzing instead the total time taken up with audience interactions (but excluding the time taken by formal discussants), the coefficient is of a similar magnitude (+1.1 minutes compared with a mean of 12.1 minutes, a rise of 9.2 percent).

Comparing across fields in columns (3) through (5), we see that this gender disparity in the number of questions is driven by the Macro programs, which exhibit a very large gender differential where women are asked 4.4 more questions than men during an average NBER talk. The other outcomes shown on Table 7 suggest that these additional questions in Macro talks are asked mostly by men, as was the case in the spring seminars sample. Moreover, the additional questions aimed at female Macro presenters start coming earlier in the talk (nearly seven minutes earlier) and at times before questions are even allowed. Some of these additional questions asked of women are deferred, which may suggests that they were not asked at an appropriate moment during the talk. In fact, the norm that questions should be held until later is 23 percentage points more likely to be breached when a woman is presenting in a Marco talk compared to a man. Finally, none of the additional questions received by women in Macro talks are rated as valuable, constructive, or collegial.

The other advantage of the NBER sample is the ability to test whether the differential treatment of presenters based on their gender can be mitigated by any specific format. Table 8 shows the same outcomes of interest, splitting the sample by program format. Surprisingly, we find that having a discussant and/or Q&A at the end does not mitigate the differential treatment of women presenters. Indeed, women receive more questions than men even in those presentations that had formal discussants. This appears to at least partly reflect audiences being less likely to respect the formatting rules when facing female presenters: they are 8 percentage points more likely to ask a question before the official question time begins. The only mitigating factor appears to be the "moratorium" on questions in the first 5 or 10 minutes of the talk: with the caveat that this represents a very small sample of presentations (N=45), we find that the moratorium completely undoes (if anything, reverses) the gender gap. And this appears to be the result of fewer "clarifying" questions that end up being deferred anyway or followed up on later when asked too early. (We should note that these formats are not randomly assigned, and so these findings may not admit a causal interpretation.)

Overall, the findings from the NBER Summer Institute results appear quite consistent with the spring seminar results presented earlier. Female presenters receive more questions that are not favorably rated, with the effects more pronounced in the Macro talks. While the results are at times imprecise, they are of roughly the same magnitude as those observed in the spring seminars where we are able better able to control for heterogeneity across audience and sub-fields. Moreover, these results shed light on the potential for different seminar formats to mitigate the disparate treatment of women during economics seminars.

6. Discussion and Conclusion

Our analysis finds notable differences between how male and female presenters are treated during economics seminars, and these effects are evident in an array of both objective and subjective indicators. Women are asked about 12 percent more questions per seminar, and they are asked more patronizing and hostile questions, and those questions are more likely to be rated as unfair. Our analysis of a major conference finds evidence of a roughly similar gender differential—and the differences may be quite a bit larger in macroeconomics—but our estimates there are less precise. These estimates are robust to a range of controls, and we obtain similar findings when analyzing data coded by people of different demographic groups.

What should we make of these results? One might respond that the differences in how women and men are treated, while notable—and in most cases, statistically significant—may not seem particularly large. Alternatively, one might note that the size of the gender gap we document is in line with the "unexplained" gender gap in wages in the United States, estimated at 8-18% (Blau and Kahn 2017). Our own view is that it is hard to know whether to call these effects "large" or "small," at least partly because it is hard to think about the long-term consequences of receiving 12 percent more interruptions throughout one's career or when on the job market. Moreover, we have focused on the measurable, rather than the unmeasurable. It might not be the magnitude of the questions that matters as much as the type or tone of the questions, particularly those that could be considered "demoralizing." Many of us have heard stories of friends and colleagues whose bad experiences in seminars have led them to re-evaluate whether a career in economics is really the best choice for them.

It seems unlikely to us that these findings reflect an explicit plan by seminar attendees to treat women differently. Instead, they may speak to implicit bias, or perhaps more darkly, an undercurrent of misogyny in a male-dominated culture (Wu 2020). As such, it seems likely that the same biases that lead women economists to be treated differently in the seminar room may also be evident in many other domains of their professional lives, perhaps shaping decisions about publication, hiring, promotions, tenure, the allocation of professional resources and so on. Indeed, one characterization of the emerging literature on gender biases within the economics profession is that every rock we look under reveals yet another way in which existing institutions are biased against women. (The exception is Donald and Hamermesh's (2006) finding that women are more likely to be elected to be officeholders of the American Economic Association—though this may itself be a tax on women's time rather than a career boost.) The cumulative effect of these various disadvantages may well be far greater than that of any individual bias considered in isolation.

Of course, some caveats are in order when generalizing our results to the entire profession. Our study is based on a convenience sample of seminars largely drawn from the top 30 economics departments in the United States, supplemented with a major elite-level conference. Seminar dynamics in other countries, other academic institutions, or other professional settings (and particularly other fields!) might operate quite differently.

The AEA Code of Professional Conduct clearly states that "Economists have both an individual responsibility for their own conduct, and a collective responsibility to promote professional conduct" by "developing institutional arrangements and a professional environment that promote free expression." Our findings suggest that the current institutional arrangements are not gender neutral. We are starting to see some evidence of recognition of this fact. The AEA Task Force on Best Practices for Professional Conduct in Economics recommend "setting and enforcing rules of responsible behavior by attendees at conference and seminar presentations" (Bayer et al. 2019).[10] Over recent months, a number of leading economics departments (including the NBER) have surveyed their members, discussed potential remedies, and set new ground rules for how they want their seminars to operate. These ground rules range from simple actions like no questions in the first ten minutes and raising one's hand to be called on by the presenter, to having a moderator who guides seminar interactions and maintains a

[10] The brochure can be downloaded here: https://www.aeaweb.org/resources/best-practices/brochure

professional environment.

A large literature in psychology and sociology has studied various ways to reduce implicit bias, especially gender bias, in the workplace—see Correll (2017) for a comprehensive overview. The main take-away of this literature is that evaluations (and this is what we do when we sit through a talk: we evaluate the claims of the speaker) are biased by gender stereotypes. Gender stereotypes still abound in the world at large, and economists are not immune to them: our own group of self-selected—hence arguably ideologically biased—coders harbor such stereotypes, as evidenced by the Implicit Association Tests results shown in Figure A1. Evidence suggests that we can mitigate the impact of implicit bias by "slowing ourselves down" (Eberhardt 2020) and going systematically through pre-defined criteria or a checklist (Correll 2017). In the context of seminars or conferences, this would mean for example taking time to ask ourselves: "How important is the answer to this question at this time?" "Could I find the answer if I looked through the paper?" "What is the likelihood that the information I am after will be provided in later slides?" "Will this question further the cause of scientific inquiry?"

We conclude by noting that we have documented a troubling gender difference within our culture, and we hope that this will stimulate further research on specific changes that economists can make to create a more inclusive, equitable and constructive environment.

REFERENCES

Allgood, Sam, Lee Badgett, Amanda Bayer, Marianne Bertrand, Sandra E. Black, Nick Bloom, and Lisa D. Cook. 2019. "AEA Professional Climate Survey: Final Report." AEA.

Bayer, Amanda, Sebnem Kalemli-Özcan, Rohini Pande, Cecilia Elena Rouse, Anthony A Smith Jr, Juan Carlos Suárez Serrato, and David W Wilcox. 2019. "Best Practices for Economists: Building a More Diverse, Inclusive, and Productive Profession." *American Economic Association*. https://www.aeaweb.org/resources/best-practices.

Blair-Loy, Mary, Laura E Rogers, Daniela Glaser, Y L Wong, Danielle Abraham, and Pamela C Cosman. 2017. "Gender in Engineering Departments: Are There Gender Differences in Interruptions of Academic Job Talks?" *Social Sciences* 6 (1): 29.

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

Boustan, Leah, and Andrew Langan. 2019. "Variation in Women's Success across PhD Programs in Economics." *Journal of Economic Perspectives* 33 (1): 23–42.

Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri. 2020. "Are Referees and Editors in Economics Gender Neutral?" *The Quarterly Journal of Economics* 135 (1): 269–327.

Carter, Alecia J, Alyssa Croft, Dieter Lukas, and Gillian M Sandstrom. 2018. "Women's Visibility in Academic Seminars: Women Ask Fewer Questions than Men." *PloS One* 13 (9): e0202743.

Chari, Anusha, and Paul Goldsmith-Pinkham. 2017. "Gender Representation in Economics Across Topics and Time: Evidence from the NBER Summer Institute." National Bureau of Economic Research Working Paper No. 23953.

Chen, Jihui, Myongjin Kim, and Qihong Liu. 2017. "Gender Gap in Tenure & Promotion: Evidence from the Economics Ph.D. Class of 2008." Unpublished. https://cpb-us-w2.wpmucdn.com/about.illinoisstate.edu/dist/0/144/files/2019/10/placement-2017-02-07.pdf.

Correll, Shelley J. 2017. "SWS 2016 Feminist Lecture: Reducing Gender Biases in Modern Workplaces: A Small Wins Approach to Organizational Change." *Gender & Society* 31 (6): 725–50.

Davenport, James R. A., Morgan Fouesneau, Erin Grand, Alex Hagen, Katja Poppenhaeger, and Laura L. Watkins. 2014. "Studying Gender in Conference Talks--Data from the 223rd Meeting of the American Astronomical Society." *ArXiv Preprint ArXiv:1403.3091*.

Doleac, Jennifer, Erin Hengel, and Elizabeth Pancotti. 2020. "Econ Seminar Diversity." 2020. https://econseminardiversity.shinyapps.io/EconSeminarDiversity/.

Donald, Stephen G, and Daniel S Hamermesh. 2006. "What Is Discrimination? Gender in the American Economic Association, 1935-2004." *American Economic Review* 96 (4): 1283–92.

Eberhardt, Jennifer L. 2020. *Biased: Uncovering the Hidden Prejudice That Shapes What We See, Think, and Do*. Penguin Books.

Ginther, Donna K, and Shulamit Kahn. 2004. "Women in Economics: Moving up or Falling off the Academic Career Ladder?" *Journal of Economic Perspectives* 18 (3): 193–214.

Ginther, Donna K, and Shulamit Kah. 2021. "Women in Academic Economics: Have We Made Progress?" Naitonal Bureau of Economic Research Working Paper No. 28743.

Hengel, Erin. 2018. "Publishing While Female. Are Women Held to Higher Standards? Evidence from Peer Review." Cambridge Working Papers in Economics. Faculty of Economics, University of Cambridge.

Hinsley, Amy, William J. Sutherland, and Alison Johnston. 2017. "Men Ask More Questions than Women at a Scientific Conference." *PloS One* 12 (10): e0185534.

Jacobi, Tonja, and Dylan Schweers. 2017. "Justice, Interrupted: The Effect of Gender, Ideology, and Seniority at Supreme Court Oral Arguments." *Virginia Law Review*. Northwestern Law & Econ Research Paper No. 17-03. https://ssrn.com/abstract=2933016.

Koffi, Marlène. 2021. "Innovative Ideas and Gender Inequality." https://www.dropbox.com/s/nvu6qzm3qolgspw/Innovative_ideas_women_mkoffi.pdf?dl=0.

Miller, Michael G., and Joseph Sutherland. 2018. "The Effect of Gender, Party and Seniority on Interruptions at Congressional Hearings." Unpublished.

Paredes, Valentina A., Daniele M. Paserman, and Francisco Pino. 2020. "Does Economics Make You Sexist?" National Bureau of Economic Research No. W27070.

Sarsons, Heather. 2017. "Recognition for Group Work: Gender Differences in Academia." *The American Economic Review* 107 (5): 141–45. https://doi.org/10.1257/aer.p20171126.

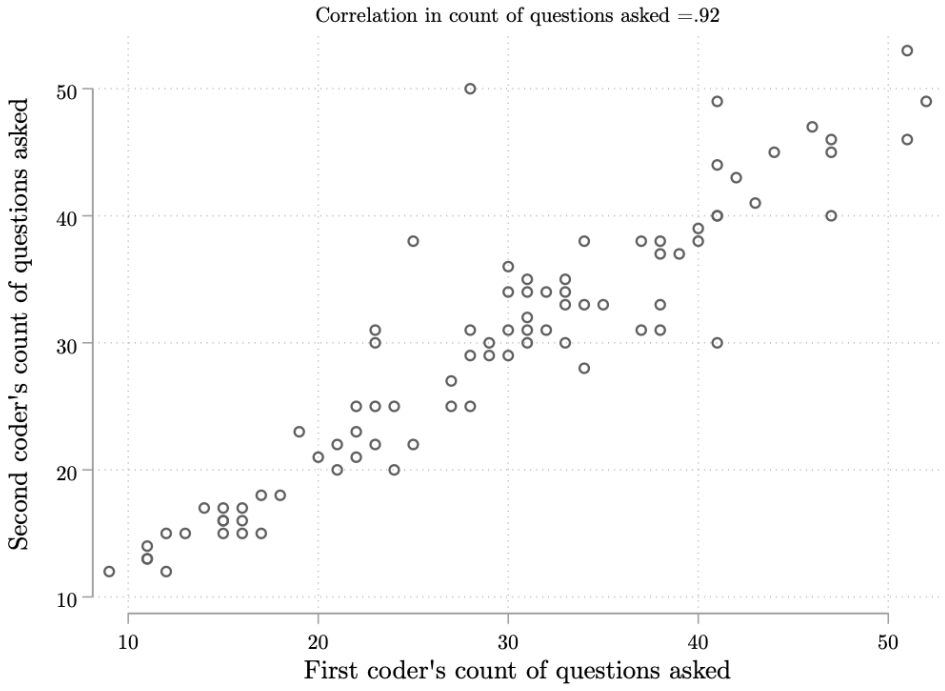
Schmidt, Sarah J., Stephanie Douglas, Natalie M. Gosnell, Philip S. Muirhead, Rachel S. Booth, James R. A. Davenport, and Gregory N. Mace. 2017. "The Role of Gender in Asking Questions at Cool Stars 18 and 19." *ArXiv Preprint ArXiv:1704.05260*.

Wu, Alice H. 2020. "Gender Bias among Professionals: An Identity-Based Interpretation." *The Review of Economics and Statistics* 102 (5): 867–80. https://doi.org/10.1162/rest_a_00877.

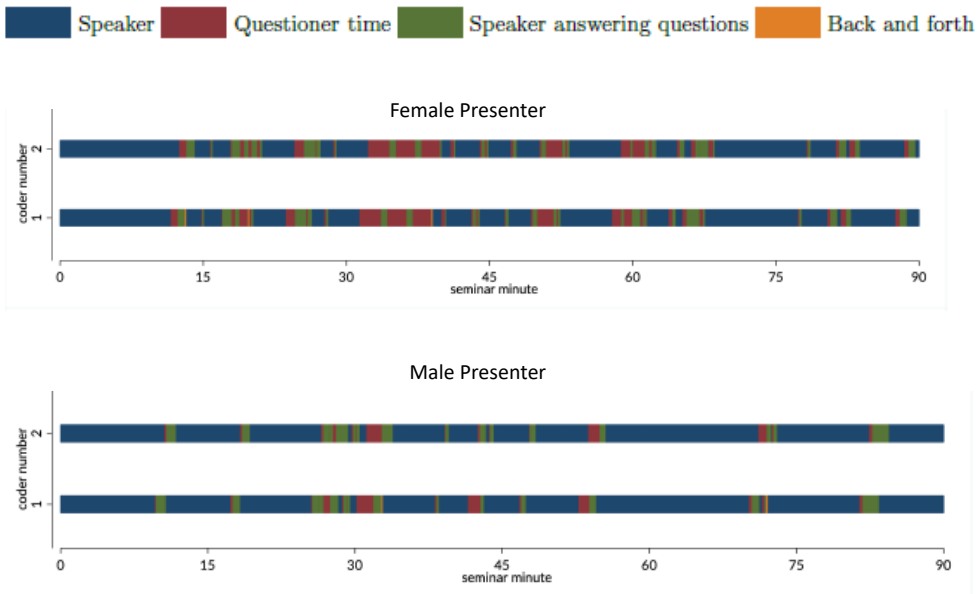
Zacchia, Giulia. 2020. "What Does It Take to Be Top Women Economists? An Analysis Using Rankings in RePEc." *Review of Political Economy*, December, 1–24. https://doi.org/10.1080/09538259.2020.1848624.

Figure 1: Coder Reliability

Panel A: Inter-Coder Reliability (N=84 talks coded by two coders)

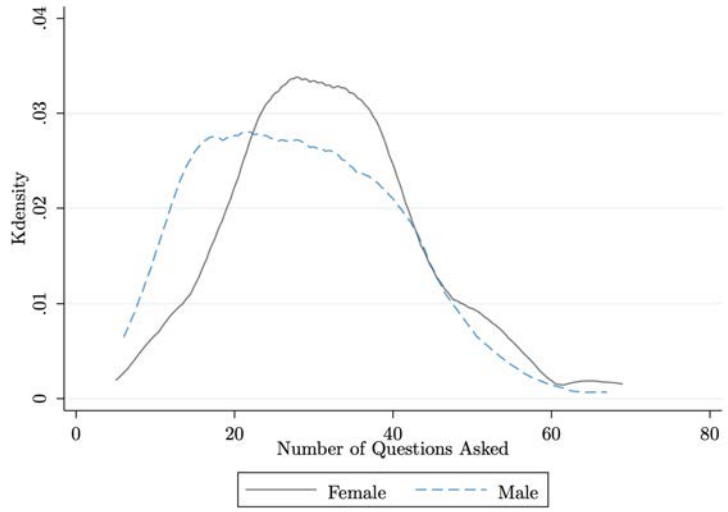


Panel B: Two Examples of Coded Seminar Timelines
(Talks Coded by Two Separate Coders)



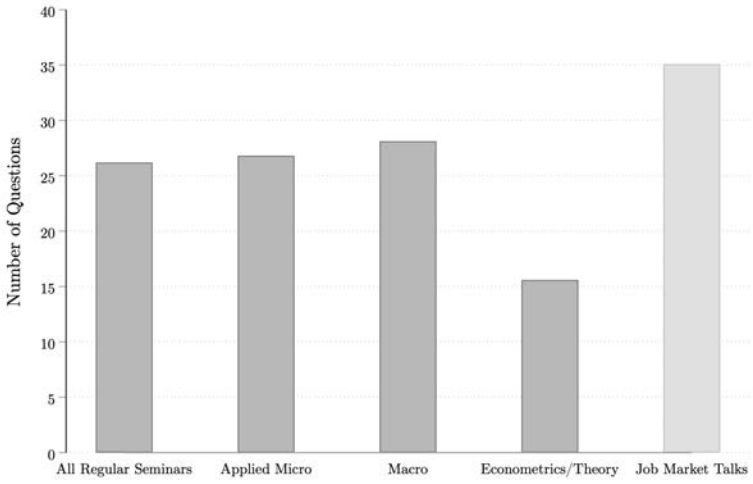
Note: Both pairs of timelines are from the same two coders, one male and one female, and the coded talks are at the same seminar series at the same institution.

Figure 2: Density of Number of Questions Asked By Gender of the Presenter



Note: Pooled sample of regular seminar talks and job maket talks N=460 talks over 576 talk-coder pairs. For the regular seminar talk sample, N=284 talk over 336 talk-coder pairs. For the JMT sample, N=176 talks over 240 talk-coder pairs. Observations are weighted by the inverse number of coders for each talk. P-value of Kolmogorov-Smirnov test of equality of the two distributions is ≤ 0.001.

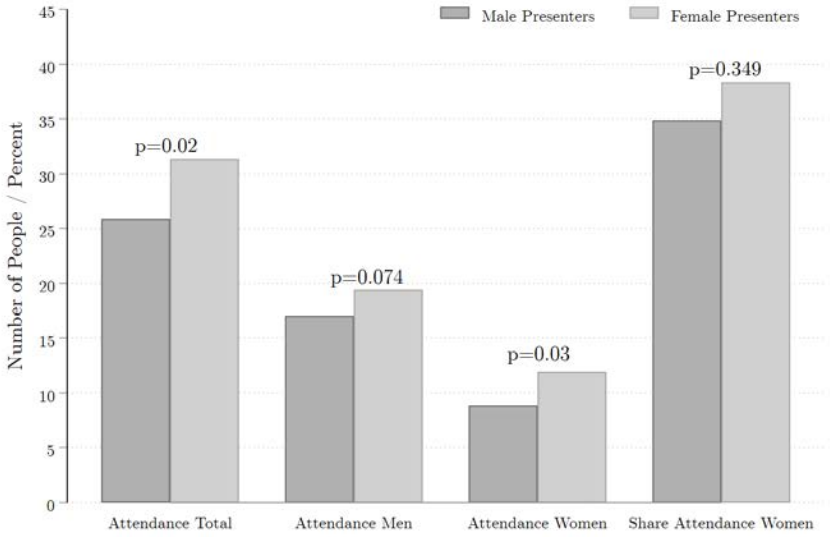
Figure 3: Number of Questions Asked in Seminar by Seminar Field/Type



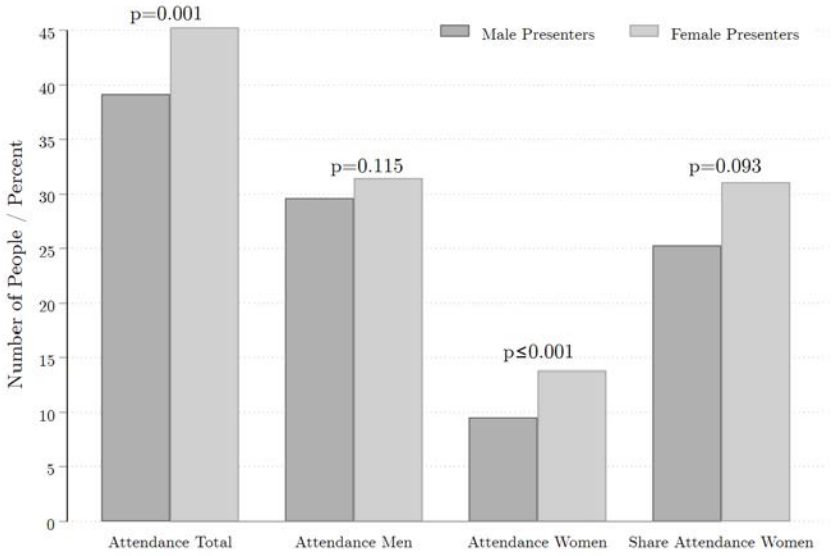
Note: For the regular seminar talk sample, N=284 talks over 336 talk-coder pairs. For the JMT sample, N=176 talks over 240 talk-coder pairs. Observations are weighted by the inverse number of coders for each talk.

Figure 4: Gender Differences in Attendance

Panel A: Regular Seminar Talks Only (N=336)



Panel B: Job Market Talks Only (N=240)



Note: For the regular seminar talk sample, N=284 talks over 336 talk-coder pairs. For the JMT sample, N=176 talks over 240 talk-coder pairs. P-values reported on top of bars show differences between male and female presenters. P-values are from regressions including seminar series and JEL code fixed effects. The y-axis indicates the number of people for the "attendance" bars and the percent female for the "share attendance" bars.

Table 1: Sample Characteristics

	Pooled Sample	**Regular Talks**	**Job Market Talks**	**T-test (JMT= Regular)**
	Mean	Mean	Mean	P-value
Speaker Characteristics				
Female Presenter	0.385	0.320	0.489	≤ 0.001
Top 10 Home Institution	0.446	0.264	0.739	≤ 0.001
Top 20 Home Institution	0.572	0.423	0.813	≤ 0.001
Talk Characteristics				
Length				
60 minutes	0.070	0.113	0.000	≤ 0.001
75 minutes	0.235	0.345	0.057	≤ 0.001
80 minutes	0.228	0.180	0.307	0.001
90 minutes	0.467	0.363	0.636	≤ 0.001
Seminar Institution				
Top 10 Seminar Institution	0.402	0.419	0.375	0.313
Top 20 Seminar Institution	0.607	0.532	0.727	≤ 0.001
Field				
Applied Micro	0.661	0.761	0.500	≤ 0.001
Macro	0.193	0.165	0.239	0.051
Econometrics/Theory	0.146	0.074	0.261	≤ 0.001
Attendance				
Total	33.095	27.611	42.143	≤ 0.001
Men	22.592	17.799	30.516	≤ 0.001
Women	10.500	9.816	11.627	0.002
Rules				
Any rules specified	0.059	0.074	0.034	0.050
No questions first 5-10 min/during intro	0.043	0.067	0.006	≤ 0.001
No questions last 5-10 min	0.015	0.007	0.028	0.113
Coder Charactersitics				
Female	0.708	0.725	0.681	0.289
Field is Applied Micro	0.704	0.678	0.746	0.097
Upper (4th-6th Year) PhD Student	0.349	0.297	0.432	0.002
Non-PhD or Unknown Coder	0.090	0.129	0.028	≤ 0.001
Observations	576	336	240	

Note: For the regular seminar talk sample, N=284 talks over 336 talk-coder pairs. For the JMT sample, N=176 talks over 240 talk-coder pairs. Observations are weighted by the inverse number of coders for each talk. Ranking for top 10 and 20 institutions is from the US News and World Report 2017 Rankings. For the regular seminar talks, fields are based on seminar series. For the job market talks, we collected field data ex-post based on the paper's content. The last column reports the p-value of the test that the means for the JMT and regular talk samples are equal.

Table 2: Interactions During Economics Seminars: Summary Statistics

	Pooled Sample		Regular Talks		Job Market Talks		T-test (JMT= Regular)
	Mean	SD	Mean	SD	Mean	SD	P-value
Quantitative Outcomes							
Number of Questions Asked							
Total	29.60	12.07	26.20	11.33	35.08	11.20	0.000
By Men	24.68	12.14	20.78	11.06	30.96	11.16	0.000
Faculty	22.86	12.65	18.47	11.46	29.92	11.21	0.000
Students	1.82	2.52	2.31	2.77	1.04	1.79	0.000
By Women	5.08	5.49	5.71	6.13	4.06	4.07	0.000
Faculty	4.51	5.04	4.89	5.54	3.90	4.06	0.015
Students	0.57	1.41	0.82	1.71	0.16	0.50	0.000
Number of Questions in First 10 Minutes	2.04	1.77	1.89	1.67	2.26	1.90	0.023
Number of Questions Answered	26.83	11.47	23.45	10.41	32.28	11.00	0.000
Number of Questions Deferred	1.41	1.85	1.07	1.71	1.96	1.94	0.000
Minutes							
Elapsed Until First Question	7.22	4.83	7.37	4.94	6.98	4.66	0.358
Taken Up by Questions	27.05	9.16	24.33	8.74	31.45	8.07	0.000
Taken Up by Questioner	10.89	5.89	8.85	4.46	14.19	6.42	0.000
Taken Up by Speaker Answers	13.15	5.31	12.62	5.63	14.01	4.65	0.003
Taken Up by Back and Forth	3.01	4.23	2.85	4.09	3.26	4.45	0.301
Share of Time Spent on Questions	0.33	0.10	0.31	0.11	0.36	0.09	0.000
Qualitative Outcomes							
Type of Question							
At least one question is a:							
Clarification	0.93	0.25	0.93	0.26	0.94	0.24	0.643
Suggestion	0.73	0.45	0.75	0.43	0.69	0.46	0.169
Comment	0.83	0.38	0.84	0.37	0.81	0.39	0.366
Criticism	0.54	0.50	0.47	0.50	0.67	0.47	0.000
Follow-Up	0.60	0.49	0.53	0.50	0.70	0.46	0.000
Number of Questions that are:							
Clarification	10.58	9.00	9.15	7.46	12.89	10.65	0.000
Suggestion	2.63	2.89	2.75	3.03	2.43	2.65	0.208
Comment	4.47	4.64	4.38	4.65	4.61	4.62	0.574
Criticism	2.14	3.43	1.70	3.31	2.84	3.51	0.000
Follow-Up	2.35	3.16	1.87	2.82	3.12	3.50	0.000
Tone of Question							
At least one question is:							
Supportive	0.42	0.49	0.42	0.49	0.42	0.49	0.914
Patronizing	0.22	0.41	0.15	0.36	0.33	0.47	0.000
Disruptive	0.17	0.37	0.12	0.33	0.24	0.43	0.001
Demeaning	0.06	0.25	0.03	0.18	0.11	0.32	0.001
Hostile	0.07	0.25	0.04	0.19	0.12	0.32	0.001
Number of Questions that are:							
Rated For Tone	3.48	6.70	3.79	7.19	2.99	5.80	0.176
Supportive	2.54	6.02	3.29	6.81	1.34	4.20	0.000
Patronizing	0.59	1.88	0.31	1.11	1.05	2.64	0.000
Disruptive	0.32	0.97	0.22	0.75	0.49	1.23	0.007
Demeaning	0.10	0.42	0.06	0.34	0.16	0.52	0.009
Hostile	0.10	0.47	0.05	0.24	0.20	0.68	0.003
Coder Rated Questions as Overall	0.07	0.26	0.05	0.23	0.11	0.31	0.032
Audience Member is Disruptive							
Any Disrupter	0.18	0.39	0.17	0.38	0.20	0.40	0.402
More than One Disrupter	0.06	0.24	0.04	0.19	0.10	0.30	0.007
At Least One Male Disrupter	0.17	0.38	0.15	0.36	0.20	0.40	0.177
At Least One Female Disrupter	0.03	0.18	0.04	0.19	0.02	0.15	0.298
Observations	576		336		240		

Note: For the regular seminar talk sample, N=284 talks over 336 talk-coder pairs. For the JMT sample, N=176 talks over 240 talk-coder pairs. Observations are weighted by the inverse number of coders for each talk. Ranking for top 10 and 20 institutions is from the US News and World Report 2017 Rankings. The last column reports the p-value of the test that the means for the JMT and regular talk samples are equal. A question can have more than one label (e.g. both a suggestion and a comment).

Table 3: Gender Differences in Number of Questions Asked during a Talk

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Estimates Full Sample					
	Pooled Sample of Regular Talks Plus JMTs					
Female Presenter	3.768	3.764	2.555	2.673	2.471	2.426
OLS SE	(1.077)***	(1.016)***	(0.915)***	(0.930)***	(0.931)***	(0.974)**
Cluster SE by Presenter	(1.235)***	(1.207)***	(1.052)**	(1.071)**	(1.039)**	(1.086)**
Cluster SE by Seminar Series	(1.068)***	(0.979)***	(0.924)***	(0.950)***	(0.958)**	(1.067)**
Cluster SE by Talk	(1.124)***	(1.067)***	(0.971)***	(0.981)***	(0.979)**	(1.035)**
Talk length (total minutes)	No	Yes	Yes	Yes	Yes	Yes
Talk series fixed effects	No	No	Yes	Yes	Yes	Yes
Presenter home institution fixed effects	No	No	No	Yes	Yes	Yes
JEL code fixed effects	No	No	No	No	Yes	Yes
Coder fixed effects	No	No	No	No	No	Yes
Number of unique talks	460	460	460	460	460	460
Number of talk-coder pairs	576	576	576	576	576	576
	Panel B: Estimates by Talk Type					
	JMTs Only		External Regular Talks Only		Internal Regular Talks Only	
Female Presenter	4.324***	3.771**	1.922	2.571**	-2.507	-0.554
	(1.415)	(1.547)	(1.233)	(1.253)	(4.448)	(3.834)
Talk length (total minutes)	Yes	Yes	Yes	Yes	Yes	Yes
Seminar series fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Presenter home institution fixed effects	No	Yes	No	Yes	No	Yes
JEL code fixed effects	No	Yes	No	Yes	No	Yes
Coder fixed effects	No	No	No	No	No	No
Number of unique talks	176	176	245	245	39	39
Number of talk-coder pairs	240	240	292	292	44	44

*** p<0.01, ** p<0.05, * p<0.1

Note: Table shows coefficient estimates from OLS regressions. The dependent variable in all columns is the total number of questions asked. All regressions in Panel A include a dummy indicating if the talk was given at the presenter's home institution ("internal talk") as well as a dummy indicating if the seminar series is a graduate student workshop (based on the seminar series' name). The groupings used for Home Institution fixed effects are shown in Table A4. The groupings used for JEL code fixed effects are shown in Table A5. Observations are weighted by the inverse number of coders for each talk. "External" regular talks refer to seminar talks where the speaker is from a different institution than the institution hosting the seminar. Robust standard errors are in parentheses.

Table 4: Gender Difference in the Type of Questions Asked and in the Number of Questions Asked and Deferred/Ignored By Gender and Seniority of the Asker

	Coefficient on Female Presenter				
	Pooled Sample of Regular Talks Plus JMTs			Regular Talks Only	JMTs Only
Dependent Variable:	(1)	(2)	(3)	(4)	(5)
Number of Questions Asked					
Questions by All	2.555***	2.471***	2.426**	1.767	3.771**
	(0.915)	(0.931)	(0.974)	(1.171)	(1.547)
Questions by Male Faculty	1.160	1.308	1.286	0.250	2.918**
	(0.836)	(0.859)	(0.911)	(1.036)	(1.467)
Questions by Female Faculty	1.317***	1.217***	1.182***	1.443***	0.967*
	(0.369)	(0.374)	(0.409)	(0.544)	(0.532)
Questions by Male Student	0.099	0.000	-0.016	-0.060	0.202
	(0.189)	(0.206)	(0.217)	(0.287)	(0.344)
Questions by Female Student	0.200*	0.213*	0.240**	0.351*	0.034
	(0.102)	(0.109)	(0.118)	(0.179)	(0.094)
Number of Questions Deferred or Ignored					
Questions by All	-0.171	-0.147	-0.069	-0.072	-0.392
	(0.202)	(0.223)	(0.236)	(0.294)	(0.335)
Questions by Males	-0.134	-0.109	-0.044	-0.067	-0.313
	(0.195)	(0.218)	(0.225)	(0.286)	(0.322)
Questions by Females	-0.014	-0.010	0.003	0.008	-0.024
	(0.049)	(0.050)	(0.056)	(0.062)	(0.094)
Number of Questions Labeled by Coder as:					
Clarification	1.483**	1.597***	1.337***	1.501*	1.646
	(0.615)	(0.617)	(0.516)	(0.787)	(1.075)
Suggestion	0.527**	0.480*	0.598**	0.438	0.752*
	(0.249)	(0.260)	(0.274)	(0.370)	(0.399)
Comment	-0.017	0.016	0.213	0.310	0.047
	(0.403)	(0.429)	(0.423)	(0.594)	(0.640)
Criticism	-0.133	-0.166	-0.312	0.093	-0.417
	(0.269)	(0.274)	(0.244)	(0.354)	(0.448)
Follow-Up	0.133	0.006	0.035	0.117	-0.198
	(0.254)	(0.269)	(0.246)	(0.322)	(0.504)
Non-labeled	0.711	0.609	0.556	-0.563	2.351*
	(0.790)	(0.796)	(0.585)	(0.997)	(1.315)
Talk length (total minutes)	Yes	Yes	Yes	Yes	Yes
Talk series fixed effects	Yes	Yes	Yes	Yes	Yes
Presenter home institution fixed effects	No	Yes	Yes	Yes	Yes
JEL code fixed effects	No	Yes	Yes	Yes	Yes
Coder fixed effects	No	No	Yes	No	No
Number of unique talks	460	460	460	284	176
Number of talk-coder pairs	576	576	576	336	240

*** p<0.01, ** p<0.05, * p<0.1

Notes: Each coefficient is from a separate regression. All regressions but those on the JMT sample only include a dummy indicating if the talk was given at the presenter's home institution ("internal talk") as well as a dummy indicating if the seminar series is a graduate student workshop (based on the seminar series' name). See Tables A4 and A5 for Home Institution and JEL codes groupings, respectively. Observations are weighted by the inverse number of coders for each talk. Standard errors are robust and in parentheses. Information on the characteristics of the person asking the question ("Male Faculty", "Female Faculty", "Male Student" and "Female Student") is imperfect (it is missing when coders did not know the seniority status of the asker, or it is counted multiple times in this table if multiple characteristics were selected for one question). Hence, the coefficients in front of "from Males" and "from Females" do not always exactly sum to the coefficient in front of "from All". Non-labeled questions are questions that were not tagged as a clarification, suggestion, comment, criticism or follow-up. A question can have more than one label (e.g. both a suggestion and a comment).

Table 5: Gender Differences in Tone of Questions Asked

	Coefficient on Female Presenter				
	Pooled Sample of Regular Talks Plus JMTs			Regular Talks Only	JMTs Only
Dependent Variable:	(1)	(2)	(3)	(4)	(5)
Number of Questions Rated by Coder As:					
Supportive	-0.150	-0.319	-0.418	-0.165	-0.030
	(0.306)	(0.356)	(0.366)	(0.557)	(0.373)
Patronizing	0.281**	0.286**	0.276*	0.308	0.284
	(0.134)	(0.138)	(0.150)	(0.211)	(0.207)
Disruptive	0.050	0.063	0.071	-0.026	0.112
	(0.105)	(0.101)	(0.110)	(0.092)	(0.196)
Demeaning	0.002	0.010	0.012	0.005	0.028
	(0.041)	(0.041)	(0.046)	(0.053)	(0.068)
Hostile	0.093**	0.080*	0.074*	0.055	0.097
	(0.045)	(0.043)	(0.039)	(0.044)	(0.086)
Patronizing, Disruptive, Demeaning or Hostile	0.410**	0.426**	0.419**	0.329	0.515*
	(0.179)	(0.179)	(0.192)	(0.257)	(0.277)
Number of Questions from **Males** Rated As:					
Supportive	-0.152	-0.280	-0.374	-0.106	-0.112
	(0.269)	(0.305)	(0.320)	(0.482)	(0.343)
Patronizing	0.189*	0.190*	0.178	0.181	0.204
	(0.104)	(0.106)	(0.115)	(0.136)	(0.187)
Disruptive	0.047	0.059	0.060	-0.045	0.135
	(0.096)	(0.093)	(0.101)	(0.081)	(0.182)
Demeaning	-0.014	-0.002	-0.000	-0.014	0.026
	(0.038)	(0.037)	(0.042)	(0.048)	(0.063)
Hostile	0.082**	0.069*	0.069*	0.052	0.081
	(0.042)	(0.041)	(0.037)	(0.041)	(0.084)
Patronizing, Disruptive, Demeaning or Hostile	0.298**	0.312**	0.300*	0.180	0.437*
	(0.146)	(0.145)	(0.156)	(0.177)	(0.258)
Number of Questions from **Females** Rated As:					
Supportive	0.002	-0.037	-0.042	-0.045	0.073
	(0.103)	(0.124)	(0.133)	(0.208)	(0.066)
Patronizing	0.093	0.097	0.100	0.130	0.080
	(0.061)	(0.065)	(0.074)	(0.114)	(0.070)
Disruptive	0.009	0.012	0.019	0.033	-0.023
	(0.026)	(0.026)	(0.029)	(0.039)	(0.036)
Demeaning	0.016	0.012	0.012	0.019	0.002
	(0.013)	(0.013)	(0.014)	(0.019)	(0.016)
Hostile	0.010	0.011	0.005	0.002	0.016
	(0.013)	(0.012)	(0.013)	(0.015)	(0.022)
Patronizing, Disruptive, Demeaning or Hostile	0.120	0.124	0.128	0.167	0.078
	(0.074)	(0.077)	(0.087)	(0.134)	(0.078)
Talk length (total minutes)	Yes	Yes	Yes	Yes	Yes
Talk series fixed effects	Yes	Yes	Yes	Yes	Yes
Presenter home institution fixed effects	No	Yes	Yes	Yes	Yes
JEL code fixed effects	No	Yes	Yes	Yes	Yes
Coder fixed effects	No	No	Yes	No	No
Number of unique talks	460	460	460	284	176
Number of talk-coder pairs	576	576	576	336	240

*** p<0.01, ** p<0.05, * p<0.1

Note: Each coefficient is from a separate regression. All regressions but those on the JMT sample include a dummy indicating if the talk was given at the presenter's home institution ("internal talk") as well as a dummy indicating if the seminar series is a graduate student workshop (based on the seminar series's name). See Tables A4 and A5 for Home Institution and JEL codes groupings, respectively. Observations are weighted by the inverse number of coders for each talk. Standard errors are robust and in parentheses.

Table 6: Summary Statistics Across Fields, NBER SI Sample

	Mean For:		Coefficient on Dummy For:	
	Full Sample	Macro Talks	Micro	Finance
Dependent Variable	(1)	(2)	(3)	(4)
Characteristics of Presenter				
Female	0.275	0.187	0.163***	0.071
			0.045	0.067
Junior	0.445	0.368	0.150***	0.028
			0.050	0.075
Student, Post-Doc, or Non-Academic Speaker	0.199	0.251	-0.093**	-0.062
			0.041	0.060
Top 20 Institution	0.526	0.503	0.044	0.014
			0.051	0.076
Meeting rules/structure				
Talk duration (min)	38.651	46.249	-10.340***	-19.878***
			1.252	1.855
Regular Seminar Format	0.357	0.386	0.044	-0.386***
			0.047	0.070
Discussant	0.490	0.374	0.126**	0.419***
			0.049	0.073
Moratorium	0.102	0.158	-0.074**	-0.158***
			0.030	0.045
Q&A Session at the End	0.508	0.386	0.086*	0.614***
			0.047	0.070
Audience				
Number of Women	17.997	12.047	11.470***	3.220*
			1.272	1.883
Number of Men	45.409	40.383	5.748**	17.177***
			2.251	3.335
Total Number of Questions Asked	14.385	17.287	-3.759***	-8.295***
			1.042	1.543
Share of questions that are neither Valuable, nor Constructive, nor Collegial	0.781	0.838	-0.118***	0.004
			0.031	0.046

*** p<0.01, ** p<0.05, * p<0.1

Note: There are 443 unique talks over 447 talk-coder pairs. Among macro talks, there are 171 unique talks over 173 talk-coder pairs. Each row is for a separate regression. N=447 for all regressions. Columns (1) and (2) report the mean of the dependent variable for all talks and only macro talks. Columns (3) and (4) report the coefficients on the dummies for Micro talks and Finance talks for regressions of the dependent variable on both dummies. Observations are weighted by the inverse number of coders for each talk. Standard errors are in parentheses.

Table 7: Gender Differences in Number of Questions Asked during Talk, NBER SI Sample

		Coefficient on Female Presenter					
	Mean (SD)	All Talks	Finance Talks	Macro Talks	Micro Talks	Talks by Senior Presenters	Talks by Presenters not from Top20 Institution
Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Total Number of Questions Asked	14.385	1.263	0.890	4.367*	-0.332	1.340	1.474
	(10.538)	(0.889)	(0.940)	(2.291)	(0.822)	(1.507)	(1.396)
Number of Questions Asked by Men	11.720	1.067	0.612	3.641	-0.166	1.286	1.174
	(9.925)	(0.820)	(0.848)	(2.215)	(0.641)	(1.379)	(1.279)
Number of Questions Deferred	3.179	0.456	1.982	0.852**	-0.047	0.916	0.225
	(3.687)	(0.311)	(1.232)	(0.397)	(0.447)	(0.627)	(0.504)
Minutes Taken up by Questions and Answers	12.145	1.117**	0.903	1.900*	0.677	1.378	1.266
	(5.516)	(0.516)	(0.955)	(1.126)	(0.632)	(0.912)	(0.820)
Minutes Elapsed Until First Question	18.243	-1.132	0.100	-6.766***	1.488	-1.618	-2.838
	(15.210)	(1.259)	(3.439)	(2.368)	(1.588)	(2.368)	(1.999)
At least one question asked before Discussant/Q&A	0.352	0.070	0.067	0.232**	0.013	0.119	0.121
	(0.479)	(0.054)	(0.105)	(0.114)	(0.077)	(0.089)	(0.089)
Number of Questions neither Valuable, nor Constructive, nor Collegial	11.874	0.824	0.013	4.895**	-1.111	0.920	1.481
	(10.836)	(0.962)	(1.089)	(2.412)	(0.939)	(1.675)	(1.504)
Share of questions that are Clarifications	0.302	-0.016	0.075	-0.050	-0.019	0.007	-0.019
	(0.265)	(0.026)	(0.065)	(0.053)	(0.032)	(0.043)	(0.041)
Share of questions that are Suggestions	0.230	0.031	-0.068	0.006	0.073**	0.028	0.052
	(0.217)	(0.022)	(0.066)	(0.039)	(0.029)	(0.042)	(0.038)
Share of questions that are Comments	0.244	-0.011	-0.052	-0.028	0.016	0.019	0.034
	(0.233)	(0.025)	(0.064)	(0.042)	(0.035)	(0.049)	(0.040)
Share of questions that are Criticisms	0.130	-0.002	-0.044	0.023	-0.007	-0.020	-0.002
	(0.150)	(0.016)	(0.049)	(0.035)	(0.018)	(0.031)	(0.029)
Share of questions that are Follow-Ups	0.099	0.002	0.053	0.019	-0.013	0.008	0.028
	(0.146)	(0.015)	(0.035)	(0.031)	(0.019)	(0.023)	(0.025)
Total Attendance	63.372	0.485	1.832	-3.336	0.508	5.113	1.662
	(28.832)	(2.594)	(7.860)	(3.695)	(3.865)	(5.089)	(3.928)
Observations	447	447	59	173	215	160	164

*** p<0.01, ** p<0.05, * p<0.1

Note: Each cell shows the coefficient estimate on "Female Presenter" from a different regression. Regressions include Field x Format fixed effects (where Format can take 5 values: regular seminar format, seminar format with moratorium, Discussant without Q&A, Discussant with Q&A, Q&A only). There are 443 unique talks over 447 talk-coder pairs. Observations are weighted by the inverse number of coders for each talk. Standard errors in parentheses.

Table 8: Gender Differences in Number of Questions Asked, by NBER Talk Format

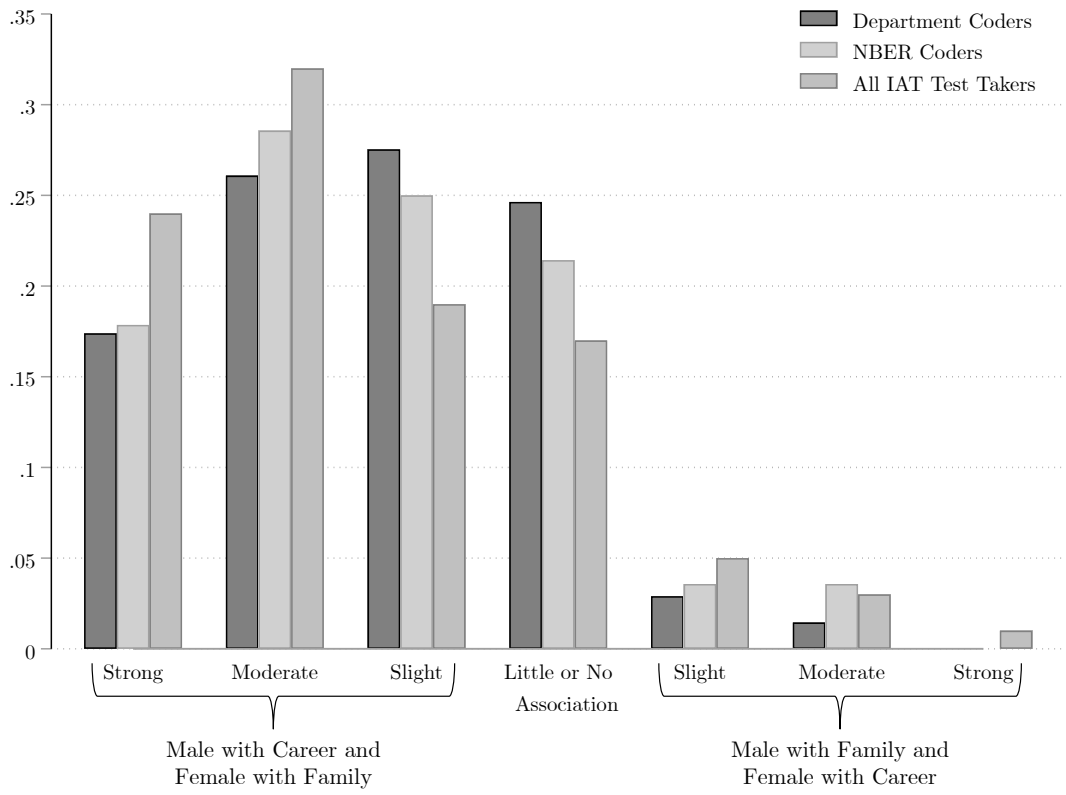
	Coefficient on Female Presenter			
	Regular Seminar Format	Talks With Discussant	Talks with Q&A at the End	Talks with Moratorium
Dependent Variable	(1)	(2)	(3)	(4)
Total Number of Questions Asked	1.404	1.661***	1.576***	-2.704
	(2.220)	(0.617)	(0.557)	(2.464)
Number of Questions Asked by Men	1.484	1.324**	1.189**	-1.732
	(2.029)	(0.533)	(0.479)	(2.375)
Number of Questions Deferred	0.128	0.907*	0.906*	-1.361*
	(0.338)	(0.521)	(0.506)	(0.778)
Minutes Taken up by Questions and Answers	0.997	1.048*	1.225**	-0.515
	(1.029)	(0.619)	(0.570)	(1.448)
Minutes Elapsed Until First Question	-0.605	-2.297	-1.492	0.562
	(1.581)	(2.052)	(2.048)	(1.514)
At least one question asked before Discussant/Q&A	N/A	0.081	0.085	N/A
		(0.060)	(0.058)	
Number of Questions neither Valuable, nor Constructive, nor Collegial	1.416	1.116	0.879	-2.767
	(2.350)	(0.746)	(0.677)	(2.681)
Share of questions that are Clarifications	-0.021	-0.004	0.007	-0.191*
	(0.045)	(0.031)	(0.033)	(0.103)
Share of questions that are Suggestions	0.023	0.033	0.042	-0.040
	(0.026)	(0.035)	(0.036)	(0.040)
Share of questions that are Comments	-0.014	0.006	0.004	-0.067
	(0.038)	(0.038)	(0.039)	(0.060)
Share of questions that are Criticisms	-0.010	0.001	0.012	-0.067
	(0.023)	(0.024)	(0.025)	(0.042)
Share of questions that are Follow-Ups	-0.019	0.034	0.044**	-0.156**
	(0.023)	(0.021)	(0.021)	(0.063)
Total Attendance	0.592	-0.496	1.682	0.726
	(4.228)	(4.032)	(3.919)	(4.509)
Observations	159	220	228	45

*** p<0.01, ** p<0.05, * p<0.1

Note: Each cell shows the coefficient estimate on "Female Presenter" from a different regression. Regressions include Field x Format fixed effects (where Format can take 5 values: regular talk format, talk format with moratorium, Discussant without Q&A, Discussant with Q&A, Q&A only). There are 443 unique talks over 447 talk-coder pairs. Observations are weighted by the inverse number of coders for each talk. Standard errors in parentheses.

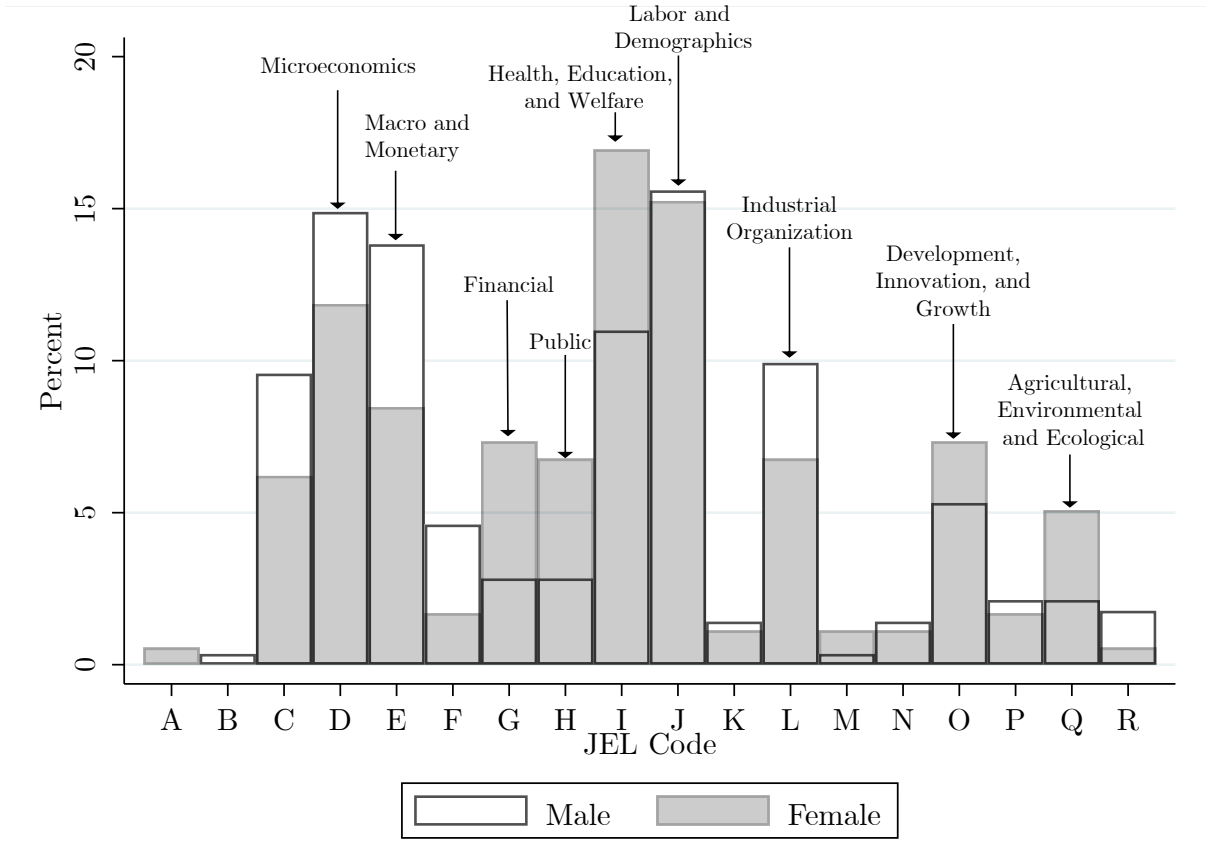
Appendix A: Appendix Figures and Tables

Figure A.1: Implicit Association Test for Gender Career Stereotypes



Source: Data provided by the Seminar Dynamics Collective and Project Implicit at Harvard University (https://implicit.harvard.edu/implicit/takeatest.html).
N=69 out of 77 Department Coders, N=28 out of 29 NBER Coders. 3 Coders are in both samples.

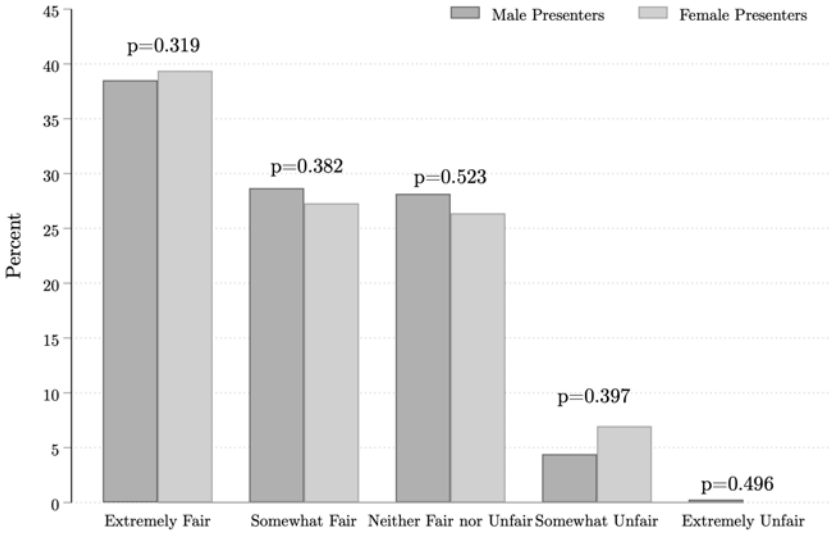
Figure A.2: Distribution of JEL Paper Codes By Gender of the Presenter



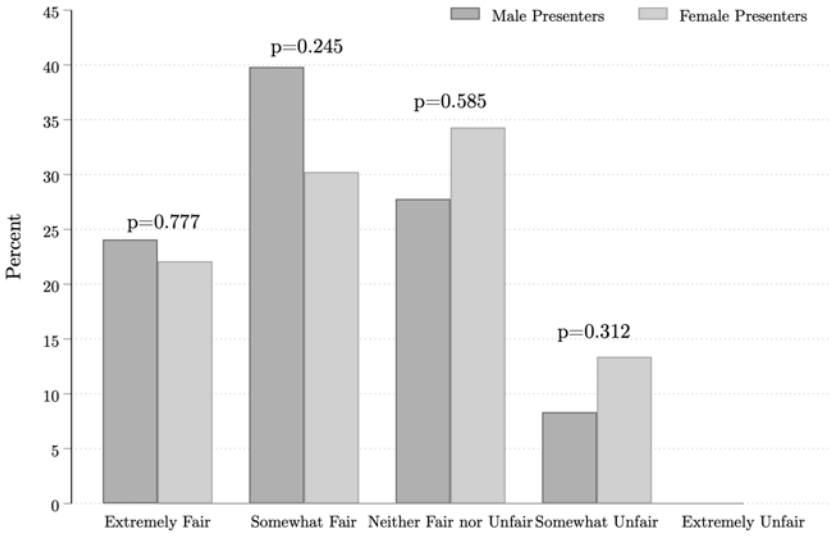
Note: The black outline represents the bar for males and the grey bar is for females. Sample is composed of N=460 seminars over 576 talk-coder pairs. Among the 460 observations, a little over half (51 percent) had JEL codes that were specified by the author on the title page of their paper. If the JEL codes listed mapped to more than one top-level (single-digit) code, then the most frequent JEL code was chosen. In the event of a tie, the most relevant JEL code was chosen. We assigned a top-level (single-digit) JEL code to the remaining observations based on the abstract. Observations are weighted by the inverse number of coders for each talk.

Figure A.3: Gender Differences in Fairness of Questions

Panel A: Regular Seminar Talks Only (N=336)



Panel B: Job Market Talks Only (N=240)



Note: For the regular seminar talk sample, N=284 talks over 336 talk-coder pairs. For the JMT sample, N=176 talks over 240 talk-coder pairs. P-values reported on top of bars show differences between male and female presenters. P-values are from regressions including seminar series and JEL code fixed effects.

Table A.1: Summary Statistics by Gender

	Female	Male	Difference (Female - Male)	
	Mean	Mean	Mean	P-value
At the Talk-Coder Level				
Seminar Field				
Applied Micro	0.441	0.488	-0.047	0.296
Macro	0.051	0.134	-0.083	0.001
Theory/Econometrics	0.023	0.060	-0.037	0.026
Job Market Talk	0.486	0.318	0.168	≤ 0.001
Seminar Institution				
Top 10 Seminar Institution	0.435	0.382	0.053	0.228
Top 20 Seminar Institution	0.638	0.587	0.052	0.244
Talk Length				
60 minutes	0.073	0.067	0.006	0.785
75 minutes	0.209	0.251	-0.042	0.266
80 minutes	0.254	0.212	0.042	0.273
90 minutes	0.463	0.470	-0.007	0.882
Attendance				
Total	38.032	30.032	8.000	≤ 0.001
Men	25.188	20.977	4.211	0.001
Women	12.829	9.055	3.774	≤ 0.001
Observations	229	347	576	
At the Speaker Level				
Home Institution				
Top 10 Home Institution	0.354	0.359	-0.005	0.932
Top 20 Home Institution	0.460	0.520	-0.060	0.300
JEL Paper Codes				
A General Economics and Teaching	0.009	0.000	0.009	0.160
B History of Economic Thought, Methodology, and Heterodox Approaches	0.000	0.004	-0.004	0.477
C Mathematical and Quantitative Methods	0.071	0.085	-0.014	0.647
D Microeconomics	0.106	0.152	-0.046	0.245
E Macroeconomics and Monetary Economics	0.071	0.135	-0.064	0.082
F International Economics	0.027	0.045	-0.018	0.413
G Financial Economics	0.053	0.031	0.022	0.331
H Public Economics	0.062	0.027	0.035	0.116
I Health, Education, and Welfare	0.168	0.117	0.052	0.191
J Labor and Demographic Economics	0.150	0.148	0.002	0.952
K Law and Economics	0.009	0.009	-0.000	0.991
L Industrial Organization	0.062	0.090	-0.028	0.378
M Business Admin and Business Econ, Marketing, Accounting, Personal Econ	0.018	0.004	0.013	0.225
N Economic History	0.018	0.018	-0.000	0.988
O Economic Development, Innovation, Technological Change, and Growth	0.106	0.054	0.052	0.079
P Economic Systems	0.027	0.027	-0.000	0.985
Q Agricultural and Natural Resource Econ, Environmental and Ecolological Econ	0.035	0.027	0.008	0.666
R Urban, Rural, Regional, Real Estate, and Transportation Econ	0.009	0.022	-0.014	0.376
Y Miscellaneous Categories	0.000	0.004	-0.004	0.477
Z Other Special Topics	0.000	0.000	0.000	.
Observations	113	223	336	

Note: For the pooled sample, N=460 talks over 576 talk-coder pairs. For the regular seminar talk sample, N=284 talks over 336 talk-coder pairs. For the job market talk sample, N=176 talks over 240 talk-coder pairs. Ranking for top 10 and 20 Institutions is from the US News and World Report 2017 Rankings. Observations are weighted by the inverse number of coders for each talk for outcomes at the talk level (Seminar Field, Seminar Institution and Talk Length outcomes).

Table A.2: Summary Statistics for Job Market Candidates by Gender

	Female	**Male**	**All**	**T-test (Female = Male)**
	Mean	Mean	Mean	P-value
Placement				
Tenure Track	0.87	0.82	0.84	0.525
Tenure Track Top10	0.19	0.16	0.17	0.732
Tenure Track Top20	0.32	0.37	0.35	0.687
Post-doc	0.03	0.04	0.04	0.847
Non-Academic Job	0.10	0.06	0.07	0.562
Placement Institution				
University	0.90	0.90	0.90	0.940
Government Agency	0.00	0.04	0.03	0.260
Private Sector	0.03	0.02	0.03	0.745
International Organization	0.06	0.00	0.03	0.073
Fed	0.00	0.02	0.01	0.430
Think Tank	0.00	0.02	0.01	0.430
Paper Publication Outcome				
Published	0.13	0.10	0.11	0.714
Published Top 5	0.06	0.06	0.06	0.953
R&R Top 5	0.16	0.12	0.14	0.628
Published or R&R Top 5	0.23	0.18	0.20	0.651
Field				
Applied Microeconomics	0.58	0.39	0.46	0.094
Econometrics	0.06	0.14	0.11	0.286
Macroeconomics	0.26	0.27	0.26	0.944
Theory	0.10	0.20	0.16	0.210
Observations	31	49	80	

Source: We coded placement and paper outcomes in December 2020 by looking at speakers' websites and online CVs. We coded Job Market Paper fields by looking at papers' abstracts.
Note: Ranking for top 10 and 20 Institutions is from the US News and World Report 2017 Rankings.

Table A.3: Google Scholar Citations : Summary Statistics Among Speakers of Regular Talks

	Female		**Male**		**All**		**T-test (Female = Male)**
	Mean (SD)	Count	Mean (SD)	Count	Mean (SD)	Count	P-value
Has a google scholar profile (%)	84	83	91	174	89	257	0.126
	(37)		(29)		(32)		
Citation count as of December 2018	4,120	70	5,685	158	5,204	228	0.255
	(8,412)		(10,010)		(9,557)		
Had citations as of December 2018 (%)	99	70	98	158	98	228	0.804
	(12)		(14)		(13)		
2018 citations/citation count as of Dec 2018 (%)	23	69	24	155	23	224	0.815
	(14)		(16)		(15)		
Observations	83		174		257		

Source: We collected citation data in April 2021 by looking at speakers' Google Scholar profiles.

Note: "2018 citations" is the speaker's number of citations in the year 2018. Only speakers who gave talks at regular seminars are included. Standard deviations in parentheses.

Table A.4: Home Institution Summary Statistics

	# of Distinct Institutions	# of Obs in Sample	% of Obs w/ Female Speaker	# of JMT Obs in Sample	% of JMT Obs w/ Female Speaker
Top 1-6 Economics Departments	6	181	0.46	128	0.52
Top 7-20 Economics Departments	15	156	0.27	66	0.30
Other US Academic Institutions	50	143	0.38	15	0.40
Other International Academic Institutions	26	80	0.56	31	0.77
Non-Academic Institutions	7	16	0.19	0	.
Total	104	576		240	

Note: Home Institution fixed effects include the dummies presented in the table above. Dummies were created for institutions for top 1-6 economics departments, for top 7-20 economics departments, for other US academic institutions, for international economics departments, and for non-academic institutions. Ranking for economics departments is from the US News and World Report 2017 Rankings. Other US Academic Institutions include top 21+ economics departments, academic institutions not ranked in the US News and World Report 2017 Rankings (e.g: no Economics PhD program offered, etc.), public policy schools and business schools.

Table A.5: JEL Code Summary Statistics

	Regular Talks		**Job Market Talks**		**Full Sample**	
	Female	Male	Female	Male	Female	Male
JEL Code Dummy: C Mathematical and Quantitative Methods	7	13	5	23	12	36
JEL Code Dummy: D Microeconomics	12	30	15	22	27	52
JEL Code Dummy: E, F or G	11	39	26	25	37	64
E Macroeconomics and Monetary Economics	7	27	9	14	16	41
F International Economics	2	9	1	5	3	14
G Financial Economics	2	3	16	6	18	9
JEL Code Dummy: H or I	23	36	32	11	55	47
H Public Economics	7	4	10	7	17	11
I Health, Education, and Welfare	16	32	22	4	38	36
JEL Code Dummy: J Labor and Demographic Economics	19	45	19	11	38	56
JEL Code Dummy: L Industrial Organization	4	14	9	21	13	35
JEL Code Dummy: Other	36	47	11	10	47	57
A General Economics and Teaching	1	0	0	0	1	0
B History of Economic Thought, Methodology, and Heterodox Approa	0	1	0	0	0	1
K Law and Economics	3	1	0	4	3	5
M Business Admin and Business Econ, Marketing, Accounting, Perso	6	2	0	0	6	2
N Economic History	5	5	0	0	5	5
O Economic Development, Innovation, Technological Change, and Gr	15	17	1	2	16	19
P Economic Systems	3	7	0	1	3	8
Q Agricultural and Natural Resource Econ, Environmental and Ecol	3	7	9	0	12	7
R Urban, Rural, Regional, Real Estate, and Transportation Econ	0	5	1	3	1	8
Observations	112	224	117	123	229	347

Note: Among the 460 observations, a little over half (51 percent) had JEL codes that were specified by the author on the title page of their paper. If the JEL codes listed mapped to more than one top-level (single-digit) code, then the most frequent JEL code was chosen. In the event of a tie, the most relevant JEL code was chosen. We assigned assigned a top-level (single-digit) JEL code to the remaining observations based on the abstract. JEL code fixed effects include dummies for JEL code C, for JEL code D, for JEL code E or F or G, for JEL code H or I, for JEL code J, for JEL code L, and one dummy for all other JEL codes.

Table A.6: Robustness Analysis: Gender Differences in Number of Questions Asked

	Pooled Sample of Regular Talks Plus JMTs			Regular Talks Only	JMTs Only
	(1)	(2)	(3)	(4)	(5)
Panel A: Poisson Specifications					
Female Presenter	2.434***	2.254***	2.195***	1.756*	3.712***
	(0.801)	(0.818)	(0.819)	(0.983)	(1.439)
Panel B: Controlling for Attendance					
Female Presenter	1.948**	2.080**	2.050**	1.588	3.271*
	(0.977)	(1.007)	(1.038)	(1.211)	(1.713)
Number of unique talks	460	460	460	284	176
Number of talk-coder pairs	576	576	576	336	240
Talk length (total minutes)	Yes	Yes	Yes	Yes	Yes
Seminar series fixed effects	Yes	Yes	Yes	Yes	Yes
Presenter home institution fixed effects	No	Yes	Yes	Yes	Yes
JEL code fixed effects	No	Yes	Yes	Yes	Yes
Coder fixed effects	No	No	Yes	No	No

*** p<0.01, ** p<0.05, * p<0.1

Note: Observations are weighted by the inverse number of coders for each talk. In Panel A, all coefficients are marginal effects from a Poisson Quasi-Maximum Likelihood Estimator (QMLE). In Panel B, all coefficients are from OLS including a control for talks' total attendance.

All regressions but those on the JMT sample only include a dummy indicating if the talk was given at the presenter's home institution ("internal talk") as well as a dummy indicating if the seminar series is a graduate student workshop (based on the seminar series's name). See Tables A3 and A4 for Home Institution and JEL codes groupings, respectively. Standard errors are robust and in parentheses.

Table A.7: Gender Differences in Number of Questions Asked, By Coder Characteristics

	Pooled Sample of Regular Talks Plus JMTs		
	(1)	(2)	(3)
Gender			
Female Presenter*Female Coder	2.436**	2.435**	2.470**
	(1.094)	(1.087)	(1.173)
Female Presenter*Male Coder	2.995*	2.640	2.387
	(1.623)	(1.677)	(1.697)
Year in Ph.D. Program			
Female Presenter*Below 4th Year Coder	2.303**	2.467**	2.418**
	(1.106)	(1.109)	(1.166)
Female Presenter*4th Year and Above Coder	4.103**	3.533**	3.598*
	(1.682)	(1.710)	(1.894)
Primary Field of Study			
Female Presenter*Applied Micro Coder	2.592**	2.617**	2.597**
	(1.036)	(1.051)	(1.040)
Female Presenter*Other Field Coder	3.466	2.982	3.027
	(2.545)	(2.605)	(2.875)
Talk length (total minutes)	Yes	Yes	Yes
Seminar series fixed effects	Yes	Yes	Yes
Presenter home institution fixed effects	No	Yes	Yes
JEL code fixed effects	No	Yes	Yes
Coder fixed effects	No	No	Yes
Number of unique talks	460	460	460
Number of talk-coder pairs	576	576	576

*** p<0.01, ** p<0.05, * p<0.1

Note: Each characteristic (e.g. gender) is tested in a separate regression. Regressions also include controls for baseline coder group characteristic (e.g. Female coder and Male coder). Regressions also include a baseline coder group for "Other or Unknown Coder" (8 observations) and the corresponding interaction coefficient "Female Presenter*Other or Unknown Code", which is not displayed in the table due to the small number of observations in this group. All regressions but those on the JMT sample only include a dummy indicating if the talk was given at the presenter's home institution ("internal talk") as well as a dummy indicating if the seminar series is a graduate student workshop (based on the seminar series's name). See Tables A3 and A4 for Home Institution and JEL codes groupings, respectively. Observations are weighted by the inverse number of coders for each talk. Standard errors are robust and in parentheses.

Table A.8: Gender Differences in Number of Questions Asked in JMT, Controlling for JMT Outcomes

	Job Market Talks Only				
	(1)	(2)	(3)	(4)	(5)
Female Presenter	4.324***	3.771**	4.213***	3.839**	4.234**
	(1.415)	(1.547)	(1.595)	(1.617)	(1.658)
Talk length (total minutes)	Yes	Yes	Yes	Yes	Yes
Seminar series fixed effects	Yes	Yes	Yes	Yes	Yes
Presenter home institution fixed effects	No	Yes	Yes	Yes	Yes
JEL code fixed effects	No	Yes	Yes	Yes	Yes
Coder fixed effects	No	No	No	No	No
Published or R&R in top 5 journal	No	No	Yes	No	Yes
Tenure track in top 10 institution	No	No	No	Yes	Yes
Number of unique talks	176	176	176	176	176
Number of talk-coder pairs	240	240	240	240	240

*** p<0.01, ** p<0.05, * p<0.1

Source: We coded placement and paper outcomes in December 2020 by looking at speakers' websites and online CVs.

Note: Each column is a separate regression. All regressions include a dummy indicating if the talk was given at the presenter's home institution ("internal talk"). See Tables A3 and A4 for Home Institution and JEL codes groupings, respectively. Top 5 Journals include Econometrica, QJE, AER, JPE and the Review of Economic Studies. Observations are weighted by the inverse number of coders for each talk. Standard errors are robust and in parentheses.

Table A.9: Gender Differences in Type of Questions Asked in Regular Talks, Controlling for Google Scholar Citations

	Regular Seminar Talks Only				
	(1)	(2)	(3)	(4)	(5)
Female Presenter	2.571**	2.663**	2.423*	2.775**	2.542*
	(1.253)	(1.325)	(1.327)	(1.370)	(1.352)
Talk length (total minutes)	Yes	Yes	Yes	Yes	Yes
Seminar series fixed effects	Yes	Yes	Yes	Yes	Yes
Presenter home institution fixed effects	Yes	Yes	Yes	Yes	Yes
JEL code fixed effects	Yes	Yes	Yes	Yes	Yes
Coder fixed effects	No	No	No	No	No
Share 2018 citations	No	Yes	No	Yes	Yes
Total citations (log)	No	No	Yes	Yes	No
Total citations (quartiles dummies)	No	No	No	No	Yes
Number of unique talks	284	284	284	284	284
Number of talk-coder pairs	336	336	336	336	336

*** p<0.01, ** p<0.05, * p<0.1

Source: We collected citation data in April 2021 by looking at speakers' Google Scholar profiles.

Note: Sample restricted to external regular talks. "Total citations" is the total number of citations as of December 2018, as reported on the speaker's google scholar profile. "Share 2018 citation" is the speaker's number of citations in 2018 divided by the speaker's total number of citations as of December 2018. All regressions include a dummy indicating if the talk was given at the presenter's home institution ("internal talk"), a dummy indicating if the seminar series is a graduate student workshop (based on the seminar series's name) and a dummy indicating if the speaker did not have a Google Schola profile as of April 2012 or any citations as of December 2017. See Tables A4 and A5 for Home Institution and JEL codes groupings, respectively. Observations are weighted by the inverse number of coders for each talk. Standard errors are robust and in parentheses.

Table A.10: Gender Differences in Number of Questions Asked, By Seminar Field

	Pooled Sample of Regular Talks Plus JMTs					
	(1)	(2)	(3)	(4)	(5)	(6)
Female Presenter*Job Market Talk	1.423	1.785	4.334***	4.569***	3.433**	3.383**
	(1.609)	(1.558)	(1.449)	(1.481)	(1.478)	(1.511)
Female Presenter*Applied Micro Seminar	2.575*	2.890**	1.590	1.777	2.438*	2.507*
	(1.360)	(1.330)	(1.248)	(1.269)	(1.275)	(1.328)
Female Presenter*Macro Seminar	5.987	4.923	-1.847	-1.936	-1.912	-1.941
	(5.853)	(5.370)	(3.930)	(3.886)	(3.909)	(4.103)
Female Presenter*Other Seminar	2.971	3.975	1.836	0.106	-0.129	-0.736
	(2.368)	(3.385)	(2.182)	(2.077)	(2.171)	(2.153)
F-test p-value	0.846	0.879	0.329	0.201	0.439	0.374
Talk length (total minutes)	No	Yes	Yes	Yes	Yes	Yes
Seminar series fixed effects	No	No	Yes	Yes	Yes	Yes
Presenter home institution fixed effects	No	No	No	Yes	Yes	Yes
JEL code fixed effects	No	No	No	No	Yes	Yes
Coder fixed effects	No	No	No	No	No	Yes
Number of unique talks	460	460	460	460	460	460
Number of talk-coder pairs	576	576	576	576	576	576

*** p<0.01, ** p<0.05, * p<0.1

Note: Fields are regular talk Applied Microeconomics, regular talk Macroeconomics, Job Market Talk and Other, which includes regular talk Theory and Econometrics. Fields are based on seminar series. There are few observations of Macroeconomics talks and only 9 observations of Macroeconomics talks given by female presenters. Regressions also include controls for baseline field group (e.g. Micro). The F-test p-value row reports the p-values of the F-test that all interaction coefficients are equal. All regressions include a dummy indicating if the talk was given at the presenter's home institution ("internal talk") as well as a dummy indicating if the seminar series is a graduate student workshop (based on the seminar series's name). See Tables A3 and A4 for Home Institution and JEL codes groupings, respectively. Observations are weighted by the inverse number of coders for each talk. Standard errors are robust and in parentheses.

Table A.11: Spring Seminars Sample Breakdown

	# of Presenters			# of Seminar Series		
	Female	Male	Total	Total	with ≥ 1 Female Presenter	with ≥ 2 Female Presenters
Job Market Talk	117	123	240	26	25	20
Regular Seminar	112	224	336	57	39	28
Regular Seminar, Applied Microeconomics	98	163	261	45	31	25
Regular Seminar, Macroeconomics	9	39	48	7	5	2
Regular Seminar, Theory	5	12	17	3	3	1
Regular Seminar, Econometrics	0	10	10	2	0	0
Internal Talk	15	31	46	13		
Graduate Student Seminar Series	6	8	14	2	2	1
Observations	229	347	576	83	64	48

Note: Fields are regular talk Applied Microeconomics, regular talk Macroeconomics, regular talk Econometrics, regular talk Theory and Job Market Talk. Fields are based on seminar series. Internal talks are talks given at the speakers' home institutions, by a faculty member or a graduate student.

Table A.12: Gender Differences in Number of Questions Asked during a Talk, By Paper JEL Code

	Pooled Sample of Regular Talks Plus JMTs			Regular Talks Only	JMTs Only
Coefficient on:	(1)	(2)	(3)	(4)	(5)
Female * JEL Code C	0.273	1.494	1.242	6.250*	-4.159
	(3.390)	(3.132)	(2.969)	(3.524)	(5.236)
Female * JEL Code D	5.702*	6.842**	6.103**	3.720	4.233
	(3.055)	(2.789)	(2.722)	(3.695)	(3.758)
Female * JEL Code E, F or G	4.354	4.010	4.208	4.408	2.800
	(2.859)	(2.645)	(2.677)	(4.407)	(3.646)
Female * JEL Code H or I	1.019	0.986	0.732	1.219	-1.151
	(2.229)	(2.045)	(2.128)	(2.803)	(2.782)
Female * JEL Code J	5.805**	4.653*	3.834	5.017	1.483
	(2.663)	(2.659)	(2.668)	(3.205)	(4.637)
Female * JEL Code L	-2.120	-1.958	-1.752	-4.766	-6.297
	(4.211)	(3.541)	(3.350)	(3.824)	(4.601)
Female * JEL Code Other	5.161**	4.014*	5.006**	5.493**	3.823
	(2.325)	(2.281)	(2.348)	(2.513)	(4.689)
F-test p-value	0.449	0.441	0.417	0.348	0.511
Talk length (total minutes)	No	Yes	Yes	Yes	Yes
Seminar series fixed effects	No	No	No	No	No
Presenter home institution fixed effects	No	No	Yes	Yes	Yes
JEL code fixed effects	Yes	Yes	Yes	Yes	Yes
Coder fixed effects	No	No	No	No	No
Number of unique talks	460	460	460	284	176
Number of talk-coder pairs	576	576	576	336	340

*** p<0.01, ** p<0.05, * p<0.1

Note: Among the 460 observations, a little over half (51 percent) had JEL codes that were specified by the author on the title page of their paper. If the JEL codes listed mapped to more than one top-level (single-digit) code, then the most frequent JEL code was chosen. In the event of a tie, the most relevant JEL code was chosen. We assigned a top-level (single-digit) JEL code to the remaining observations based on the abstract. See Table A5 for composition of JEL Code "Other". The F-test p-value row reports the p-values of the F-test that all interaction coefficients are equal. All regressions but those on the JMT sample only include a dummy indicating if the talk was given at the presenter's home institution ("internal talk") as well as a dummy indicating if the seminar series is a graduate student workshop (based on the seminar series' name). See Table A4 for Home Institution groupings. Observations are weighted by the inverse number of coders for each talk. Standard errors are robust and in parentheses.

Table A.13: Other Outcomes of Interest

	Coefficient on Female Presenter				
	Pooled Sample of Regular Talks Plus JMTs			Regular Talks Only	JMTs Only
Dependent Variable:	(1)	(2)	(3)	(4)	(5)
Time Spent on Questions					
Share of Time Spent on Questions	0.010	0.011	0.011	0.020	0.001
	(0.008)	(0.008)	(0.009)	(0.012)	(0.012)
Share of Time Spent on Asking Questions	0.007	0.008	0.007	0.009	0.001
	(0.005)	(0.005)	(0.005)	(0.006)	(0.009)
Share of Time Spent on Answering Questions	0.006	0.006	0.004	0.007	0.011
	(0.005)	(0.006)	(0.006)	(0.008)	(0.008)
Share of Time Spent on Back and Forth	-0.003	-0.003	-0.000	0.005	-0.011**
	(0.004)	(0.004)	(0.004)	(0.006)	(0.005)
Timing of Questions					
Minutes Elapsed Until First Question	0.072	0.228	0.192	0.869	-0.407
	(0.428)	(0.431)	(0.451)	(0.611)	(0.595)
Number of Questions in the First 10 minutes	0.049	0.016	0.002	-0.250	0.189
	(0.163)	(0.166)	(0.181)	(0.216)	(0.255)
Particularly Disruptive Audience Members					
Any Disruptive Audience Members	0.041	0.047	0.032	0.082	-0.007
	(0.037)	(0.039)	(0.039)	(0.051)	(0.068)
Any Disruptive Audience Male Members	0.040	0.043	0.028	0.076	-0.007
	(0.037)	(0.039)	(0.039)	(0.051)	(0.068)
Any Disruptive Audience Female Members	-0.023	-0.024	-0.027	-0.020	-0.033
	(0.015)	(0.015)	(0.017)	(0.022)	(0.022)
Any Disruptive Audience Female Members	-0.023	-0.024	-0.027	-0.020	-0.033
	(0.015)	(0.015)	(0.017)	(0.022)	(0.022)
Talk length (total minutes)	Yes	Yes	Yes	Yes	Yes
Talk series fixed effects	Yes	Yes	Yes	Yes	Yes
Presenter home institution fixed effects	No	Yes	Yes	Yes	Yes
JEL code fixed effects	No	Yes	Yes	Yes	Yes
Coder fixed effects	No	No	Yes	No	No
Number of unique talks	460	460	460	284	176
Number of talk-coder pairs	576	576	576	336	240

*** p<0.01, ** p<0.05, * p<0.1

Note: Each coefficient is from a separate regression. All regressions but those on the JMT sample include a dummy indicating if the talk was given at the presenter's home institution ("internal talk") as well as a dummy indicating if the seminar series is a graduate student workshop (based on the seminar series's name). See Tables A4 and A5 for Home Institution and JEL codes groupings, respectively. Observations are weighted by the inverse number of coders for each talk. Standard errors are robust and in parentheses.

Table A.14: Gender Differences in Type of Questions Asked by Gender of the Asker

	Coefficient on Female Presenter				
	Pooled Sample of Regular Talks Plus JMTs			Regular Talks Only	JMTs Only
Dependent Variable:	(1)	(2)	(3)	(4)	(5)
Number of Questions That Are:					
Non-Labeled by Males	0.453	0.387	0.261	-0.973	2.346*
	(0.685)	(0.703)	(0.518)	(0.833)	(1.205)
Non-Labeled by Females	0.258	0.222	0.295	0.410	0.005
	(0.258)	(0.253)	(0.258)	(0.389)	(0.316)
Clarification by Males	0.764	0.915*	0.696	0.844	0.844
	(0.524)	(0.541)	(0.460)	(0.629)	(0.993)
Clarification by Females	0.732***	0.702***	0.653***	0.657*	0.840***
	(0.246)	(0.233)	(0.238)	(0.363)	(0.308)
Suggestion by Males	0.232	0.223	0.345	0.163	0.542
	(0.221)	(0.230)	(0.244)	(0.328)	(0.346)
Suggestion by Females	0.287***	0.248***	0.244***	0.272*	0.198*
	(0.086)	(0.091)	(0.092)	(0.141)	(0.108)
Comment by Males	-0.200	-0.167	0.019	0.085	-0.090
	(0.359)	(0.382)	(0.389)	(0.515)	(0.582)
Comment by Females	0.190*	0.189*	0.187	0.214	0.169
	(0.111)	(0.115)	(0.121)	(0.186)	(0.121)
Criticism by Males	-0.169	-0.214	-0.329	-0.013	-0.385
	(0.232)	(0.235)	(0.217)	(0.276)	(0.418)
Criticism by Females	0.039	0.050	0.025	0.114	-0.035
	(0.087)	(0.091)	(0.086)	(0.138)	(0.113)
Follow-Up by Males	0.060	-0.038	-0.013	0.079	-0.190
	(0.228)	(0.245)	(0.227)	(0.270)	(0.477)
Follow-Up by Females	0.084	0.058	0.062	0.053	0.022
	(0.075)	(0.079)	(0.082)	(0.112)	(0.118)
Talk length (total minutes)	Yes	Yes	Yes	Yes	Yes
Talk series fixed effects	Yes	Yes	Yes	Yes	Yes
Presenter home institution fixed effects	No	Yes	Yes	Yes	Yes
JEL code fixed effects	No	Yes	Yes	Yes	Yes
Coder fixed effects	No	No	Yes	No	No
Number of unique talks	460	460	460	284	176
Number of talk-coder pairs	576	576	576	336	240

*** p<0.01, ** p<0.05, * p<0.1

Note: Each coefficient is from a separate regression. Information on the characteristics of the person asking the question ("Male Faculty", "Female Faculty", "Male Student" and "Female Student") is imperfect (it is missing when coders did not know the seniority status of the asker, or it is counted multiple times in this table if multiple characteristics were selected for one question). Non-labeled questions are questions that were not tagged as a clarification, suggestion, comment, criticism or follow-up. All regressions but those on the JMT sample include a dummy indicating if the talk was given at the presenter's home institution ("internal talk") as well as a dummy indicating if the seminar series is a graduate student workshop. See Tables A4 and A5 for Home Institution and JEL codes groupings, respectively. Observations are weighted by the inverse number of coders for each talk. Standard errors are robust and in parentheses.

Table A.15: NBER Coder Characteristics

	Mean	SD	Number
Female	0.517	0.509	15
Field is Applied Micro	0.828	0.384	24
Upper (4th-6th Year) PhD Student	0.310	0.471	9
Observations	29		

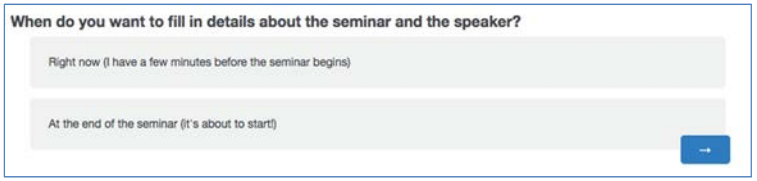
Appendix B: Qualtrics Tool User Guide

We provide below the user guide for the qualtrics tool used by coders for the university seminar sample. The tool for the NBER summer institute meetings was identical except for the General Information section and the options for the "tone" of the questions, as discussed in the main text.

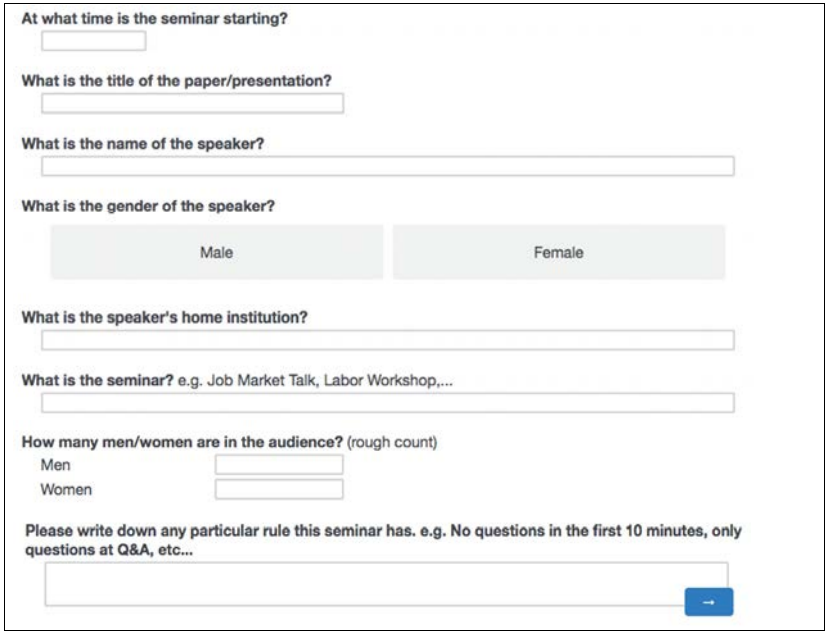
Economics Seminar Dynamics - Qualtrics Tool User Guide

A. Survey Start: General Information

After some general information, the first question you will be asked is:

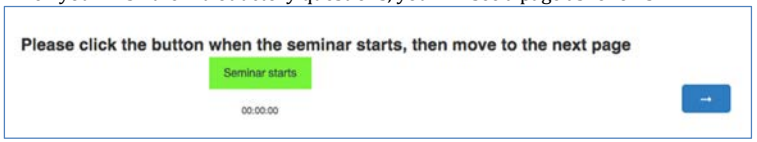


➔ If you do not have time right now, then you will get those questions at the end of the survey.
➔ If you answer right now, you will get to the following screen:

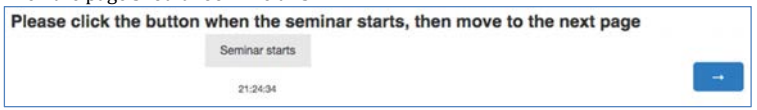


These questions should be self-explanatory. Do the best you can to be precise, please, but it is okay to approximate if you cannot count exactly how many people are in the room.

When you finish the introductory questions, you will see a page as follows:



When the seminar starts click the green button: You should click this after the speaker introductions. The seminar formally begins once the speaker takes over the floor. (Note: You may have to play a little with the mouse, only the top part of the "button" is clickable). Then the page should look like this:



I clicked the "Seminar Starts" button at 9:24 pm (21:24 for Europeans ;-). Then click the blue arrow button to get to the next page.[1]

[1] NOTE: you can click the blue button even if you didn't click the "Seminar starts" button. While we prefer you click the "Seminar starts" button for precise timing, don't worry if you forgot, we will then use the time you indicated on the first page as to when the seminar starts.

B. Recording Timing Data on Questions during the Seminar

Here is a template Question page. Each page has two panels.

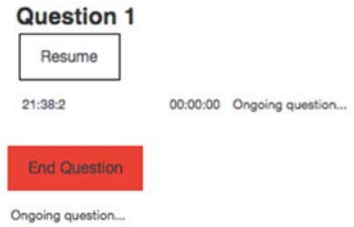
As a general rule, use the left panel only, and when you are done, click the blue arrow (12) to go to the next page that, again, will have a left and a right panel. The right panel is a shade darker, so that we don't get confused which button belongs where, while using all these buttons…

What's the idea for the right panel? The right panel is here in case one question follows another quickly, and you're still filling out stuff on the left panel, and had no time to finish and click the blue arrow button. It's almost like an "emergency" left panel. In general, you use the right panel just like the left panel. But don't make it a habit to always use it. The reason is that if you use the right panel even if you had ample time to move to the next page, then, in case there are two quick questions in succession, you won't have the right panel there to help you not lose track! You need not worry you'll run out of "Question" pages, we have plenty of those.

Now, let's move to how the Question panel is best filled out. I'll refer to the buttons through their red number in the Figure above.

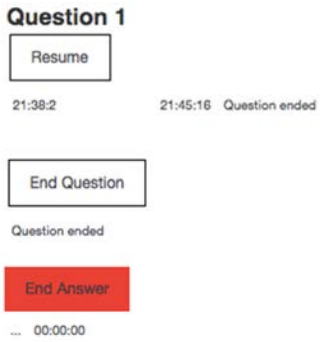
When a question is asked, click the green "Start" button.

Then two things happen as shown in the picture below:
1) The "Start" button (1) changes to a "Resume" button and turns White: Here, I clicked the button at 9:38 pm and 2 seconds.
2) The second change is the "End Question" Button (5) in the middle turns red. This alerts you that you are still in the Ongoing Question mode. Until you click End Question, we think the person who asked the question is still speaking (recall, the tool is collecting data both on how many questions there are, and how long those questions last).



When the question is over, click the Red "End Question" button.

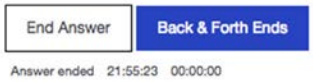
Then three things happen as shown in the picture below:
1) Below and next to the "Start" button (1), you see a time stamp and a "Question ended" text. Here, I clicked the button at 9:45 pm and 16 seconds (so the question looks like it lasted about 7 minutes).
2) The second change is the "End Question" Button (5) at the bottom turns white, and the text below turns to question ended.
3) The "End Answer" Button (9) at the bottom turns red. This alerts you that you are still in the Ongoing Answer mode, that is, until you click End Answer, we think the speaker is still answering the question. We set it up this way because we want to know how much time the speaker takes out of their seminar to answer a question. *We understand that this can feel somewhat subjective; simply use your best judgment when you think the answer is over.*



When the answer is over, click the Red "End Answer" button.

Then two things happen as shown in the picture below:
1) the "End Answer" Button (9) at the bottom turns white. This alerts you that the answer is over.
2) A new button appears next to the "End Answer" button, namely the blue "Back & Forth Ends" button. We will address this below.

Here, I clicked the end answer button at 9:55 pm and 23 s, so, about a 10 min answer.



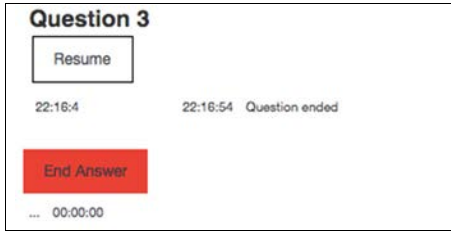
Before we go to the other buttons, and how to use them, let's consider some difficult cases on how people ask questions and answer them, let's call them Problem Cases.

C. Timing of Questions: Problem Cases

1. The question ended, but then, suddenly, the question kept going.

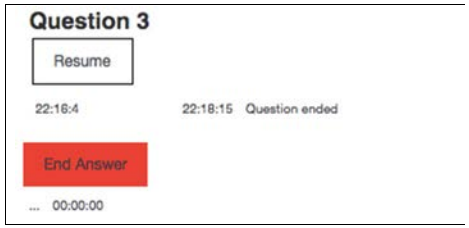
So, suppose you clicked the "End Question" button, but then the question kept going. This can also happen if you click the "End Question" button by mistake.

No worries: Simply click the "Resume" button.



Here I clicked the "Start" Question button at 22:16 and 4 seconds, and the "End Question" button 22:16 and 54 seconds: a very short 50 second question.

But then I noticed my mistake and clicked the resume button, and the "Question ended" text changes to "Ongoing question…" and the "End Question" button turns Red again to alert me to the fact that once more the question is ongoing. Furthermore, the "End Answer" button once more turns white, as we are back to the question mode rather than the answer mode.



Then I clicked the "End Question" button, here at 10:18 pm, and 15 s, a much more reasonable 2 min question. And apart from the "End Question" button turning white, the "End Answer" button turns red again.

2. Back and Forth

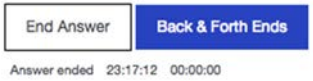
Sometimes there is not just one quick question, but there is a whole back and forth between the questioner and the speaker (potentially including several members of the audience who jump in on the same point)

If the person who asked the initial question follows up immediately with the speaker (or interrupts the speaker's response), then you can click the "Resume" button. When the speaker starts answering, click the "End Question" button again. You can do this repeatedly, so if there is a back and forth for a while you can keep hitting "resume" , "end question", "resume", "end question". At the very end of the interaction you can hit "End Answer". You do not need to hit the **Back & Forth Ends** in that case (though it's OK if you do for safety).

Backup! If you are totally overwhelmed by the speed of question and answer so that the buttons simply seem too difficult to all fill out, or there are too many different audience members interjecting, this is where the **"Back & Forth Ends"** button kicks in.

Here, I started the Question at 23:16:59 (pm) (so almost 11:17 pm) and ended is just a few seconds later.

Then I clicked the "End Answer" button at 23:17:12, so a little more than a minute after the question ended.



Then I was overwhelmed by the rapid question and answer back and forth, that at the end of this, I clicked the "Back & Forth Ends" button, at 23:20:58, so, almost another 3 minutes later. The "Back & Forth Ends" button keeps being blue, since, in case it still goes on, you simply click it, when you think it is really over, again.

And indeed, it went on, and I simply clicked it again when I thought it was over, now at 23 minutes and 23 seconds, so, about 2 minutes later.

And still ongoing, so I simply click it again when I think it is over, now at 23 hours, 24 min and 50s.

Since you can click the blue "Back & Forth Ends" Button as often as you need, you should feel free to press it as soon as you think the back and forth is over, since you can always come back and change your mind by clicking it again when it is now (maybe truly) over, etc. So, This should help you to not overestimate how long the back and forth takes. If you're not sure it's over, simply assume it is, and if it's not, simply click it again!

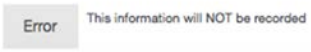
3. Someone quickly asks another question before I managed to fill out the whole left panel about the question and answer:

This is what the right panel is for, you can click the "Start" button (3) on the right panel, and your left panel is still "alive" that is you can still fill out whatever you need while the next question is happily under way on the right panel.

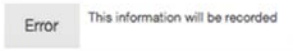
4. The whole Question was a mistake, there never was a question…

No worries, this is why we have the "Error" button at the very end.

When you click it, it will turn into:



So, the information will not be recorded.
But wait, wait… I clicked the error button by mistake!!! No worries, simply click it again, and voila, it turns into the following:

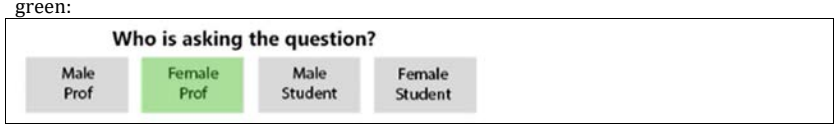


And so on and so forth….

D. Recording Specifics on Questions

You can enter information about the question during the question, answer, or after both have finished, but only before moving to the next page. Once you get to the next page, you can NOT return to the previous page.

How to use those other buttons:
Here I thought a Female Prof was asking the question, so, I click that button and it turns green:



You can technically click as many buttons as you want, please only click one. We realize that at times you may not be sure who someone, but just use your best judgment. Classify post-docs as students, and visiting faculty as profs.

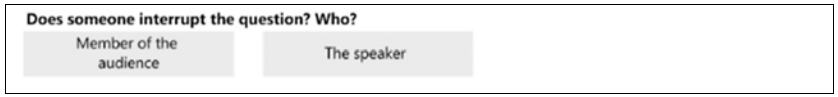
Next is the:

You should think of this as assessing the interaction towards the speaker. We aren't asking you to code the intention of the person making the comment, nor how it was taken by the speaker. But rather, it's your assessment of the tenor of the comment in a scientific setting. You may leave this blank for many interjections, only filling this up if you think it is warranted. *You can click as many buttons as you deem appropriate, so, an interaction can be, e.g. just supportive, just patronizing or both. Or nothing. Use your judgment.*

The options are:

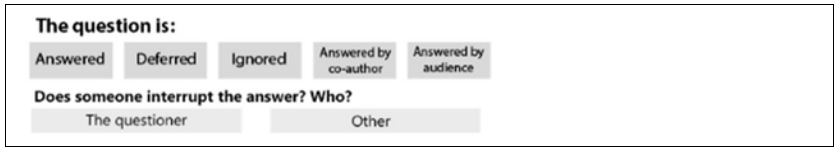
- **Supportive:** For example, I provide the speaker with a great example they can use. Or provide an answer to a problem. Or I tell them why I find their insight useful.
- **Patronizing:** A comment that may be apparently kind or helpful, but betrays a feeling or sense of superiority over the speaker. A comment could be both supportive and patronizing if the interjection acts as if the speaker can't answer themselves.
- **Disruptive:** Here we think of interactions that disrupt the flow of the seminar, maybe shifting the talk into a completely different direction, away from the speaker and their resarch.
- **Demeaning:** A comment that – in some measure – causes the speaker to lose their dignity or the respect of others. A demeaning comment is less about the scientific point being made, and more about shifting the focus to the speaker and undermining their status as an expert.
- **Hostile:** A comment that is unnecessarily antagonistic, aggressive, confrontational or combative. Hostility describes an aggressive interaction, one that you may not want to encounter as a speaker. Hostility is not required to make a scientific point.

Next is the:



This is pretty straightforward.

Then is the:



Again, pretty straightforward, and once more you can click as many as you feel apply.

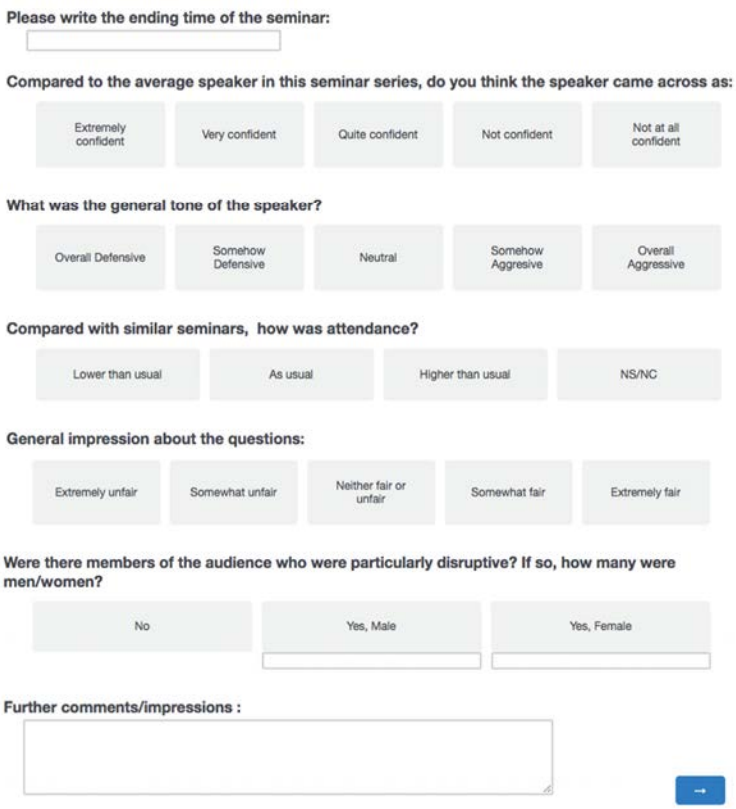
E. Coding Q&A Sessions

Some seminar series have a Q&A sessions at the end. To deal with this, do as follows:

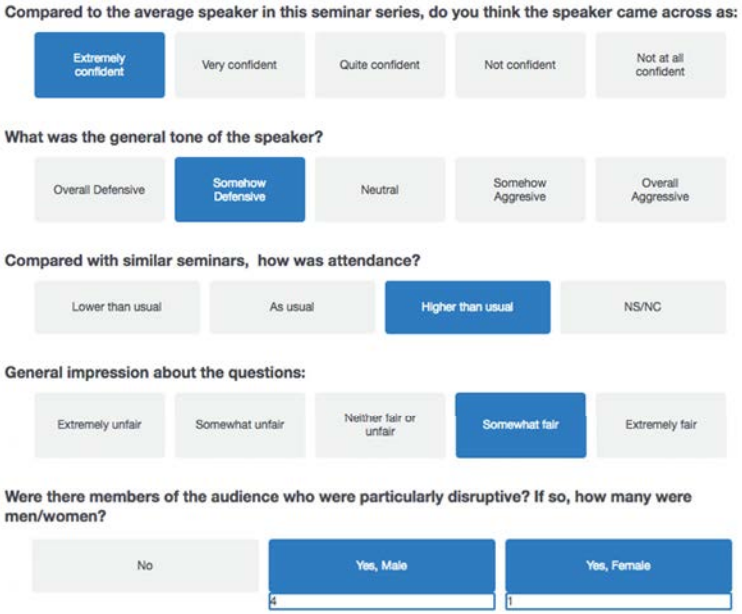
1. For the first question during the Q&A session, please write "Start of Q&A session" in the comment box.
2. Code questions as normal (as instructed above).

F. Finally, the seminar is over, no more questions:

Simply click the blue Arrow button 3 times and you get to the following page:



You can use your local time for the beginning and end of a seminar. Here I filled it out randomly for demonstration purposes. The last two boxes indicate that 4 men and 1 woman were disruptive (it was a very disruptive imaginary talk).



Note: If you did not do so before the seminar, you may then get to a page that enables you to enter demographic information about the seminar (talk title, etc.)

Then click the blue Arrow button once more, and you are DONE!

Thank you!

Appendix C: Coder Comments and Impressions

We provide here an exhaustive list of comments and impressions left by coders after finishing encoding a talk. At the end of the Qualtrics tool, there was also an open-ended comment box for coders to note any further comments or impressions. Comments were left for 123 talks, 72 of which included impressions and are reported below. Quotes were left untouched except for spelling mistakes which were corrected and seminar institution or speaker which were de-identified. Comments or portion of comments about coding, attendance, or the Qualtrics tool itself, are excluded from the list below. No other changes were made.

	Speaker Gender	Comment
1.	F	"A certain male professor was especially disruptive, making comments constantly."
2.	F	"A few people were antagonistic at the beginning of the seminar, until a male faculty member asked for no further questions until she got past the introduction of her talk (around minute 40)."
3.	F	"definitely a lot more questions than usual, probably because the subject matter is relatively provocative"
4.	F	"Disruptive member of the audience wasn't aggressive or unfair, but did interrupt often"
5.	M	"He was somehow disruptive because his interventions tend to be very long."
6.	M	"I found that a lot of the questions were about the background industry; the economic related questions were fair I thought."
7.	M	"I think the quantity & length of questions was far too high although no individual question was unfair."
8.	F	"I think the questions were generally fair but there were too many questions and her responses were long. That combined with a high number of slides left her to not finish."
9.	F	"I would not characterize the questions as unfair, but they were very challenging and skeptical, largely deservedly. However, they also got a little out of hand, redundant, excessive, etc."
10.	M	"I wouldn't say this went well but he was confident and unflappable through a series of hard questions. Questioning wasn't unfair by and large but a few questions that were asked perhaps too early and would have been answered later in the talk anyway"
11.	F	"In the first few minutes the presenter made the comment that no one had asked her questions and a professor in the audience explained that we as a community are trying to actively improve seminar culture; in particular, allowing the person to get through the introduction before being peppered with questions. The seminar went much longer than it should've partly because the presenter was given a couple extra minutes but then also because all 4 of the male professors left in the room asked a question after the official end time."

12.	M	"lots of questions because it was a theory paper but with a lot of non-theory people in the audience. Speaker was not always immediately clear in his answers so lot of back and forth in the questions"
13.	F	"lots of faculty talking amongst self as speaker was still presenting, particularly in the last 8 minutes"
14.	F	"Lots of nice suggestive comments"
15.	F	"Lots of questions but just to make sure people understand"
16.	F	"Lots of questions but the audience was very jovial/friendly"
17.	F	"Many of the same question repeated in different ways"
18.	M	"many questions from people of other fields"
19.	F	"Many repeated questions on same issue, talk was partially derailed. Felt speaker had good charisma but she became a bit discouraged toward end (kept her cool though)"
20.	M	"More clarifying questions than comments and suggestion on the content."
21.	M	"No one's questions were unfair, but one person in particular asked far too many"
22..	M	"One male prof had questions that were partially him being funny/smart and partially questions. I usually marked as disruptive in those cases."
23.	M	"One male professor asked questions past the point of usefulness because he was annoyed at the speaker's framing in two places in the talk."
24.	M	"One professor asked a lot of clarifying questions which in general I feel are fine but after a certain point I feel like were disruptive because there were like 5-6 of these."
25.	F	"Overall the seminar had a neutral tone. The speaker received many questions, and some criticisms were raised, but the speaker was not treated badly by the audience. As usual in JM talks, most of the questions come from male profs, but about 20 members of the audience were profs and only 2 of them were female."
26.	F	"Presenter finished with 10 minutes to spare and opened the floor to questions. This speaker was very good at deflecting unproductive lines of questioning."
27.	F	"Quasi- flyout for speaker"
28.	M	"Questions about related papers dominated the time"
29.	F	"Questions were disruptive at the beginning, and the presenter answered each with perhaps too long of explanations, eating into her time significantly. Despite warning the room that she was running out of time, the questions continued. Nearing the end, one male professor insisted on an answer to a previous question with which he was unsatisfied, continued to speak over her for a time when she tried to move on, and instigated an entire corner of the room to talk over her. There was no time left at the end for Q&A, and despite cheery responses and confidence throughout interruptions, this closing 'question' (disruption) seemed especially demoralizing."
30.	M	"Questions were dominated by one person, but they are one of the few people that know macro and this was a macro seminar which we don't usually have. Don't think he was particularly disruptive (this person always asks a lot of questions)"
31.	M	"Seemed as if there was a lot more net picking of research than other seminars."

32.	M	"Senior professor, almost a talk rather than a seminar."
33.	M	"Several loud side conversations. Male prof arrived 15 min late, left 10 min early"
34.	F	"She is a very good presenter."
35.	F	"Somewhat more 'direct criticisms' than usual, i.e. straight out saying that they were skeptical of something."
36.	F	"Sort of an overall hostile tone, I'm not really sure why"
37.	M	"Speaker at about 1:05 asked people to hold off questions for 15 minutes so he could push through to the results"
38.	M	"Speaker encouraged questions and even used quotes. making the concepts very tangible for the younger population (graduate students)"
39.	M	"Speaker struggled, but questions where very fair and warranted based on the paper"
40.	M	"Speaker was acknowledging many of the limitations of the study."
41.	M	"Speaker was confident but did not seem particularly well-prepared to answer particular questions; I would say he was somewhere between 'not confident' and 'quite confident'." With regards to disruptions: a male professor asked a question and then had a long, whispered conversation with the female professor as the speaker continued to talk, which seemed disruptive."
42.	F	"The 2 or 3 'disruptive' audience members were not necessarily rude but would politely interrupt the speaker to ask multiple follow-ups to their questions. As a result, many of the questions are actually a series of back-and-forths that span more than one subject. Most of the questions were relevant, but a large proportion during the empirical sections seemed to veer off subject, questioning why the paper chose to answer X question instead of Y question, rather than focusing on the design or the data, hence why I viewed the questions as 'somewhat unfair'. Note on the speaker's ability: I put her as 'very confident' because she started out very confident, but due to the volume/length of questions, wound up rushing at the end. As a result, she could be considered in the 'quite confident' box - in fact, she mentioned to a professor at the end of the talk that she was not used to giving job talks in the evening (since this went until [X] pm), and that this was her worst talk so far. I honestly thought she was a strong speaker, though."
43.	M	"The audience took over the exposition for about 1 minute, and then they said 'let the author finish' that was OK."
44.	F	"the crowd was a bit crazy today"
45.	F	"The discussion was lively but with a positive tone aimed at helping the speaker improve the paper"
46.	M	"The majority of comments with many back and forth came from one professor. They were all on topic and weren't rude so I wouldn't call him disruptive, but he did monopolize time."
47.	F	"The male professor that was particular disruptive has been in all macro seminars (including this one), he is a visiting professor, and seems to be driving the seminars in a very disruptive way."

48.	F	"The speaker often ended answers with an upward inflection, making them sound almost like a question rather than an answer. She also fairly regularly gave short answers to the surface question without offering a significant response to the underlying question/concern."
49.	M	"The speaker was presenting a coauthored paper and several times (definitely more compared to other presenters in this workshop) had said something along the lines of 'I don't usually do empirical work' or , 'I don't usually spot the errors in [coauthor]'s work' when people were pointing out potential areas to improve, criticisms, etc."
50.	M	"The speaker was presenting a paper coauthored by one of the professors attending the seminar, so in general there was a more relaxed environment than usual - two generalized laughs. There were several questions, but it seemed they were coming from the engagement of all participants, and they seemed to help moving the presentation rather than stopping the speaker."
51.	F	"The speaker was reasonably confident but generally very quiet, so it was a bit difficult to hear her at times (I was sitting in the back of the room)."
52.	M	"The speaker was reasonably confident but overall, the quality of presentation was not as good as is typical for these seminars."
53.	M	"The speaker was running short on time and explicitly said he wouldn't answer questions in last 10 min."
54.	F	"The tone of the seminar was generally neutral. There were many questions and many back and forth, but they were not aggressive nor out of context with respect to the talk."
55.	M	"There was a lot of back and forth. Many questions and answers were interrupted during the talk. Some questions were partially answered by audience members before the speaker had a chance to start answering."
56.	M	"There was one older male teacher who constantly criticized the structure respectfully in which the speaker made a joke to move on, validating his opinion but dismissing it as just a comment."
57.	M	"There was one person who constantly asked very basic questions and tended to ask them too early, otherwise they would have been answered, which made his questions a bit unnecessary."
58.	F	"there was one point where 4 professors were talking and trying to clarify a point."
59.	F	"There were many times in which more people intervened, with the purpose of helping the speaker improving the paper."
60.	M	"This job market talk had arguably a lot more 'interruptions,' but these might more accurately be classified as a conversation involving the entire room. Interruptions generally flowed naturally and for the most part did not strike me as rude. None were particularly hostile though there were a number of faculty members who seemed to take issue with a lot of the paper being presented."

61. F "This presenter drew a very different 'vibe' than other job market talks this year. Many audience members asking questions apologized for asking, instructed the speaker to defer the question if she would get to it later, and generally 'pushed' less on questions. She was very polished and quick to defer questions; she kept a constant eye on the clock, particularly during questions."
62. F "This seminar was quite 'calm'. Questions did not have significant digressions, nor were they out of topic. The tone was never aggressive and even in those cases in which there were a couple of back and forth questions it was more of a conversation rather than a confrontational exchange. There were only 6 women out of 40 people attending, and only 3 female professors (none of them strongly related to the field)."
63. M "This speaker generally interrupted questions, he would not let the questioner completely finish the question."
64. M "This was a business school job talk. Turnout in IO workshop is typically predominantly male, today's attendance was 13 men, 3 women. Most of the questions/comments were made by 6 men and 1 woman."
65. M "This was a pretty disruptive and rambunctious seminar"
66. F "Two female professors were saying valid things, but it took time away from the seminar, especially since the speaker was just standing there."
67. M "Two tenured male professors were very supportive of this candidate and set the tone for the talk. There were many questions but none that were rude or overly critical."
68. M "Various times a few professors were audibly talking over the speaker, I think all male."
69. M "Very disruptive prof asked incessant methodological questions that were not helpful."
70. F "Very good presentation, answered questions very clearly"
71. F "Very low-key seminar, very well-paced, no timing issues at all. Many fewer students than usual because we are on winter break, but if anything, this led to more student questions than usual."
72. M "Very rude questions near the end of the presentation. Telling the speaker he was wrong, interrupting his concluding remarks with disagreement."