Title
Fear of Backlash: A Novel Explanation for Gender Differences in Participation in Question-and-Answer Sessions in Academic Settings

Author
Jarvis, Shoshana

Publication Date
2022

Peer reviewed|Thesis/dissertation

Fear of Backlash: A Novel Explanation for Gender Differences in Participation in Question-and-Answer Sessions in Academic Settings

by

Shoshana Nicole Jarvis

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Business

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Laura J. Kray, Chair

Professor Cameron P. Anderson

Professor Serena Chen

Spring 2022

Fear of Backlash: A Novel Explanation for Gender Differences in Participation in Question-and-Answer Sessions in Academic Settings

By

Shoshana Nicole Jarvis

Abstract

Fear of Backlash: A Novel Explanation for Gender Differences in Participation in Question-and-Answer Sessions in Academic Settings
by
Shoshana Nicole Jarvis
Doctor of Philosophy in Business
University of California, Berkeley
Professor Laura Kray, Chair

Men historically and presently hold more power and privilege than women in academia. Men are paid more than women, and their research contributions are valued more than work by women. This power differential also translates to the question-and-answer (Q&A) space, with men participating more than women. The purpose of this dissertation is to investigate gender differences with greater granularity than past research by testing when conference attendees participate and how they use their time. I also investigate psychological correlates (anxiety and fears of backlash) that could explain the gendered patterns of participation. The dissertation ends with an investigation of the influence of structural factors of how Q&A sessions are formatted on gendered participation and as potential mitigating factors for disproportionate participation.

Chapter 1 includes an initial investigation into gender differences in participation in question-and-answer sessions and feelings about participating in Q&A sessions. After an exploratory investigation with a small live conference, I found men disproportionately initiate Q&A interactions relative to the gender base rates in confirmatory analyses with a larger conference. However, there were no observed gender differences in other observed behaviors like the amount of time per question, number of points made, or challenging the speaker. In self-reports, women reported feeling less comfortable asking questions, making comments, or approaching speakers and greater fears of backlash. Men were more likely to say that they hold back asking questions to make space for others, whereas women were more likely to say that they hold back asking questions due to anxiety. Reasons why participants held back questions were replicated in a second sample.

Chapter 2 tested the explanatory role of fears of backlash in a series of vignette studies and experiments. Men expressed greater question-asking intentions, and women expressed greater fears of backlash across all studies. Fears of backlash significantly explained variance in the relationship between gender and question-asking intentions. Gender differences in fears of backlash were diminished when the audience was described as friendly. However, friendly audiences emboldened men to feel even more comfortable asking questions, maintaining the gender gap in question-asking intentions. In a sample of academics, women's fears of backlash were more impacted by the framing of the questions they would ask as critical or constructive than by the audience as critical or friendly. An ongoing study testing the effects of the format of Q&A sessions (e.g., chats vs. live) has not found gender effects on question-asking intentions or fears of backlash by format in the data collected thus far.

Chapter 3 evaluated the impact of various virtual Q&A formats on disproportionate participation by gender using field data from four conferences. In the three conferences where I had estimates for gender base rates, men took disproportionately more of the total time and word space compared to what would be expected by the gender base rates. This effect persisted when women were in the numeric majority of the conference, when the Q&A was only live video, and

when the Q&A was only through chats. Disproportionate participation did not significantly change between live and virtual formats within the same conference one year apart. In an all-chat format, gender disparities in chat text length decreased at higher status levels. In a mixed chat and video virtual conference, there was some evidence that women participated more in chats than in videos. However this cannot be compared to the gender base rates at the conference. Overall, there was greater support that divisional characteristics and particular dynamics within the sessions impacted gender disparities in participation as opposed to reliable effects due to differences in the structures of the Q&A sessions.

Overall, I replicated the effect that men disproportionately participate in live Q&A sessions, but that that is where the gender differences in participation seem to end. Other behavioral differences may be driven by the base rates of men participating more than by men being more likely to do them. Women reported feeling less comfortable and having greater fears of backlash, which were largely influenced by how they perceived the community around them. Virtual conferences continued to show disproportionate participation in both video and chat formats. The option for chat messages could potentially be a mitigating element, but the evidence supporting it is weak. The influence of the perception of the immediate audience seems to have a greater influence on gendered participation in Q&A sessions. Across studies and modalities, men’s disproportionate participation and women’s discomfort in Q&A situations was pervasive and stable. More work is necessary to determine how to improve the inclusivity of academic spaces for women.

Table of Contents

Acknowledgements

Grad school has been an adventure to say the least. I would not be here without the love and support of my village. I am so thankful to everyone below and countless other friends, lab mates, and advisors who helped me and lifted me up over the last eight years of my academic experience. A very big thank you to my committee for their advice and feedback and for supporting me to the end of my degree. I am thankful to everyone who has been with me on this journey, and I would like to highlight the support of the people below.

I would like to thank my high school math teacher, Mr. McGuirk, for your support across my high school career. I did not get the best grades, but you had faith in my abilities and pushed me to continue in advanced classes. Many others would have heard my doubts and suggested I move down a level. From this experience, I had no idea that the stereotypes about women being bad at math were not merely a relic of the past until I got to college. My math self-efficacy was preserved, allowing me to not only excel, but thrive in the statistics side of psychology.

I am incredibly grateful to my friends who were genuinely happy for me in my successes and held space for me through my many periods of hardships. In the game of academia where so much success is zero-sum, I would expect this type of support to be hard to come by. Yet, I was blessed with a community spanning multiple institutions who consistently were there for me, including but not limited to: Adriana Germano, Amanda Wang, Derek Brown, Emily Liquin, Emily Willroth, Felicia Zerwas, Gold Okafor, Kristin Donnelly, Lauren Feldman, Mark Babin, Micah Brush, Michelle Lee, Sonya Mishra, Stephen Antonoplis, Ted Schwaba, Tim Carsel, V de la Vega, Valerie McGraw, and Zachary Heinemann. I would also like to thank Mike O'Donnell who forged a path, truly saw me, and brought me with him. I am thankful to my college friends for sticking by my side throughout this rollercoaster: Lindsey Dennis and Kathryn Friason. And, thank you to Christine Sandall and Nicole Delgadillo, who always pick up the phone when I call.

Research, at its core, is and should be a team effort. I would like to thank all of the research assistants who I had the privilege of working with, who assisted me on the work in this dissertation and other projects. Being able to teach you and share the joy of learning about the world has been one of the highlights of my graduate and broader academic career.

In the same vein, I was very fortunate to have a team of women who I met early in my journey who taught me, supported me, and helped me be the best researcher I could be. They paved the way and brought me with them. Thank you to Simine Vazire, Lexi Suppes, Rebecca Hetey, and Joy McClure. And, an extra big thank you to the best academic big sister I could have ever asked for. Liz Page-Gould, you made me feel so welcome at the beginning of grad school, and your unconditional support has meant everything to me every moment since.

Thank you to my undergraduate advisor, Jeremy Miller. I didn't really know what research was when I walked into your office and asked if I could work with you. The opportunity to design my own research and manage it from start to finish was formative, and that experience gave me the foundation for managing all of my future research projects. And, your continued support across my positions has been grounding.

A very big thank you to Shannon Brady for being the absolute best of all time. I am so lucky to have been able to work with you so early in my career and for you to have stayed in my corner all these years. You taught me how to be a great mentor and manage teams and projects through leading by example. Working with you and writing with you has been a complete joy. I am the researcher I am today because of your dedication and care.

I would like to thank my primary advisor, Laura Kray. Thank you for taking me in, giving me the space to grow, and being there to catch me when I fell. Thank you for believing me, believing in me, and for trusting my judgment. The last year and a half has been a breath of fresh air. I wish we could have had more time to fully explore what we could do together, but I'm thankful for the time we were able to share.

Charlie, I could not have ever anticipated how much of an impact sitting next to you that day at that conference would have on not just my research program, but on my entire life. None of the research in this dissertation would exist without you. Thank you for empowering me to find my voice and for reminding me why I got into research in the first place. Thank you for being a fun research partner and a steadfast friend. I don't know where I would be without you.

Thank you to the patience and warmth of my family. Thank you to my brothers for listening to me talk about the minutiae of my research and academic dramas for hours. I have loved being able to have you on this journey with me. Lastly, thank you to my parents for the unconditional love and support through all the ups and downs. I could not have made it through this degree, or this pandemic, without you.

Introduction

Pervasive gender inequities persist in academia in which women are underrepresented as faculty and their intellectual contributions less valued (Ceci & Williams, 2011; Knobloch-Westerwick et al., 2013). These inequities are reflective of historical disparities in access to higher education and pervasive stereotypes about who belongs in academia (Baker & Vélez, 1996; Leslie et al., 2015). One manifestation of persistent gender-based hierarchies in academia is within participation in question and answer (Q&A) sessions, in which control of the collective discussion space corresponds to the gender hierarchies (i.e., primarily consumed by men).

Across three chapters, I explore how gender differences manifest in question-asking behavior, the structural factors that ameliorate or exacerbate gendered participation, and the factors of Q&A sessions that most impact women's anxiety, fears of backlash, and intentions to participate. In Chapter 1, I establish gender differences in Q&A sessions by testing both behavioral differences (frequency and style of participation) at Q&A sessions as well as differences in psychological perspectives toward Q&A sessions. First, I test for disproportionate participation in Q&A sessions in two live, recorded, interdisciplinary conferences. Then, I test gender differences in comfort participating in Q&A sessions and gender differences in reasons why conference attendees choose not to participate in Q&A sessions. In Chapter 2, I test the role of fears of backlash on question-asking intentions across 5 surveys and experiments. In Chapter 3, I capitalize on the natural variation in structures of virtual Q&A sessions to test which, if any, structures reduce gender disparities in conference Q&A participation.

Q&A sessions following research talks provide key opportunities for the audience to engage in scientific discourse. To the extent men engage more than women in Q&A sessions, men may continue to have more influence over the direction of science. Witnessing imbalanced use of shared dialogue space could impact women's sense of belonging in academia and be related to women withdrawing from academia (Biggs et al., 2018). The purpose of this dissertation is to evaluate the extent to which disproportionate participation by gender occurs in Q&A sessions and determine some of the causes of disproportionate participation.

Chapter 1: Live Conferences

Introduction

Participating in question and answer (Q&A) sessions is a ubiquitous aspect of the academic experience, providing opportunities for presenters to solicit feedback on their work from peers and experts in their field. As time is limited and few people can ask questions, those who do ask questions are highly visible, communicating information about who receives recognition and authority and who belongs in the field (McCluney & Rabelo, 2019; Simpson & Lewis, 2005). To the extent that diverse voices are included in the feedback process, the quality of the work also improves (Woolley et al., 2010). In academia, women's work continues to be underrepresented and undervalued (Knobloch-Westerwick et al., 2013; Vásárhelyi et al., 2021). The purpose of this research is to test for gender differences in question-asking behavior, and consequently, gender differences in attitudes towards participating in Q&A sessions.

Gender, Power, and Nonverbals

When people are in power, they use that power to display dominant behaviors and disproportionately occupy space. Having power is associated with positive affect and disinhibited behavior whereas being in low power positions is associated with attention to threat and inhibited social behavior (Keltner et al., 2003). People in power are more likely to be first, to spend more time talking, and to deviate from social norms and expectations in group settings (Burgoon & Dunbar, 2006; Galinsky et al., 2003; Hall et al., 2005).

Historically, men have held more power than women, and power dynamics fall along gendered lines, particularly when it comes to using collective speaking space (Lips & Lips, 1991). Men are more likely to be perceived as dominant after speaking longer compared to women (Mast, 2002). Women are more likely to experience backlash for self-promotional acts because they violate gender norms (Rudman, 1998) and are more likely to avoid male-dominated domains due the anticipated lack of power they would have in those spaces (Chen & Moons, 2015). Asking questions and taking collective space could be construed as self-promotional because it is a public display of knowledge. Women's anxiety around participating may be warranted because they are more likely to experience backlash for speaking (Brescoll, 2011; Dupas et al., 2021).

Gender and Power Differentials in Academia

Within academic spaces, men hold more power than women, and their work is taken more seriously than work produced by women. Female professors are underrepresented among professors, paid less, and less likely to be promoted across fields including psychology, STEM fields, and medicine (Ceci & Williams, 2011; Gruber et al., 2021; Kaatz & Carnes, 2014; Shen, 2013). Women are more likely to have innovative dissertations but less likely to have their novel contributions lead to successful careers (Hofstra et al., 2020). Work by women is less valued than work by men, and they are less likely to be cited, less likely to be invited to be colloquium speakers, and more likely to be assigned non-promotable tasks in academia relative to gender base rates (Babcock et al., 2017; Knobloch-Westerwick et al., 2013; Larivière et al., 2013; Nittrouer et al., 2018; Schroeder et al., 2013).

The effects of gender bias also begin early in the academic pipeline. Female prospective graduate students are less likely to get responses from faculty members after requests for a meeting in a week compared to male prospective graduate students (Milkman et al., 2012). Faculty members show preference towards male students and offer them higher salaries and more mentorship (Moss-Racusin et al., 2012). And even once women make it through the door, as fields include more women, they are less likely to be included in collaboration networks (Wang et al., 2019). With men having historically more access to academic space and being more represented in higher power positions, it would be plausible for them to reproduce these power dynamics with their behaviors in Q&A sessions.

Chilly Climates and Threat

Women are perceptive of their environments and use their perceptions to make judgments about whether they belong in those spaces (Murphy et al., 2007). Environmental cues can range from aspects of the physical space to numerical representation of women in groups and directly relate to women's sense of belonging, feelings of identity threat, and interest in continuing to invest in the space (Cheryan et al., 2009; Murphy et al., 2007). For professors, numerical representation of women in their department was related to their perceptions of gender equity and feelings of exclusion in their department (Maranto & Griffin, 2011). In turn, experiencing identity threat is related to lower cognitive bandwidth, worse performance, and concerns about belonging with meaningful impacts on retention and success in academic spaces (Cook et al., 2012; Leslie et al., 2015; Schmader, 2010; Steele, 1997; Walton & Cohen, 2007). To the extent women perceive unwelcoming environments, there are direct relationships with performance and involvement in communities.

Norms and Masculine Defaults

In addition to the climate, the norms by which we create science and engage with science has direct implications as to who becomes involved in the discourse. In a study of collaboration networks of researchers engaged in the science reform movement, the network of reproducibility researchers was predominantly male and was characterized by more insular collaboration networks. This collaboration pattern is indicative of highly valuing competition and keeping resources sequestered. In contrast, the open-science network had greater representation of women and more dispersed and collaborative networks (Murphy et al., 2020). While causality cannot be attributed to this one study, it is indicative of a relationship between norms in research practices and inclusivity in science.

Likewise, gendered participation in Q&A sessions could be reflective of how masculine defaults have been imbued into the space. Masculine defaults are structures in which more masculine attributes (e.g., individualism, competitiveness, aggression) are rewarded (Cheryan & Markus, 2020). When masculine defaults remain as the status quo, women must either try to fit a masculine mold or be left behind. This type of dynamic has been observed in conferences in which lower numerical representation of women was related to women being more likely to think they needed to behave more like men to fit in the space (Biggs et al., 2018). To extend the framework of masculine defaults to Q&A sessions, attendees must decide that they have a question and that their question is good enough for the audience to hear. Attendees are rewarded to the extent they can make this decision quickly and take initiative to control the collective space.

Gender and Q&A Sessions

While past metascience research has found that men participate in Q&A sessions at conferences more than women across STEM fields including astronomy (Pritchard et al., 2014; Schmidt & Davenport, 2017), biology (Hinsley et al., 2017), biogenetics (Käfer et al., 2018; Telis et al., 2019), and in departmental seminars in psychology and biology (Carter et al., 2018), less is known about the cause of these gender disparities, where differences emerge within question-asking behavior, and the psychological mechanisms contributing to gender disparities in question-asking behavior.

Past research has suggested that women are underrepresented among question-askers due to internal reasons such as lacking self-confidence (Carter et al., 2018). The timing in which questions are asked might matter in that women are less likely to ask the first question but have more equitable representation later on in the question asking period (Pritchard et al., 2014). Gender disparities in question-asking are attenuated when women are first to ask a question, perhaps because it invites other women to speak (Carter et al., 2018). Though men are overrepresented among senior professors, in one study, younger and older men out-questioned women in their age cohort at the same rate (Hinsley et al., 2017). While these studies provide some preliminary evidence of gender disparities in Q&A session participation, more research is necessary to better understand the types of participation, both in terms of content and style, that exhibit gender differences and their psychological correlates.

In addition to the choice of whether to participate in Q&A sessions, there could be gender differences in how men and women communicate when at the microphone, due to differences in power impacting communication styles (Burgoon & Dunbar, 2006; Tannen, 1993). Men are more likely to increase their speaking time relative to their power (Brescoll, 2011; Mast, 2002), which could manifest in men taking longer speaking turns compared to women when at the microphone. Additionally, women are more likely to adapt their speaking style to men to be more indirect (Bowles & Flynn, 2010), which could have consequences for the use of questions or comments. Lastly, men are more likely to interrupt others in conversation (Anderson & Leaper, 1998), displaying more dominance within their communication styles. A common running joke within academic spaces involves noting when askers start their remarks with, "This is more of a comment than a question…," indicating that how someone uses their time at the microphone is salient to the audience. It would be informative to know if gender differences emerge not only in whether individuals speak but also in how they use their time. Power can be performed in a multitude of ways and knowing the scope of gender differences in Q&A session behavior would inform what steps are necessary to create more inclusive spaces.

Present Studies

The present investigation builds on previous research demonstrating that men participate more in conferences by testing *how* conference attendees differ in question-asking behavior (e.g., in the types of remarks and speaking time) by gender and *why* gender differences in participation emerge. In Study 1, a small, single-track, interdisciplinary conference was used for exploratory analyses and hypothesis generation. I then preregistered my key analyses and applied them to the conferences used in the following studies. In Study 2, I coded Q&A sessions from a filmed, single-track, interdisciplinary conference for markers of participation, whether the idea presented was a question or a comment, how many ideas were presented, and markers of power/dominance in how the attendee asked the question. Study 3 tested for gender differences in comfort participating in Q&A sessions through self-reports of conference attendees from a large multi-track conference. I also tested for mechanisms of why gender differences in participating in Q&A sessions emerge by qualitatively analyzing open responses for why conference attendees hold back questions. I then tested the robustness of the mechanistic effects by replicating the analyses in an existing self-report dataset on seminar Q&A sessions. All preregistrations, materials, codebooks, data, and analysis code are available on the Open Science Framework (preregistration: https://osf.io/akceu; main project page: https://osf.io/uzmb9).

Study 1: Exploratory Conference

Methods

Conference Characteristics. All exploration occurred using data from a small, single-track, single-day, interdisciplinary conference. There were 92 attendees (59 men, 31 women, 1 nonbinary person, 1 undisclosed) from 12 fields (see Table 1). Across 11 talks and panels, there were 51 Q&A interactions. Videos from the in-person conference were posted online and available for public viewing, and a list of conference attendees was provided by the conference organizers.

Table 1. Conference attendees by discipline.

Field	Components	percent
Data Science	Data science, statistics, information sciences	10%
Economics	Economics, behavioral economics	24%
Engineering & Math	Chemical engineering, civil engineering, mathematics	5%
Law		1%
Medical Sciences	Medicine, epidemiology, radiology, pathology	4%
Metascience & Philosophy	Metascience, meta-research, philosophy, Philosophy of science, science of science	1%
Political Science		5%
Psychology and Brain Sciences	Psychology, neuroscience, behavioral marketing, micro management, cognitive science, social psychology	12%
Science Communication	Journalists, freelance writers	1%
Social Sciences	Public health, public policy, health policy, social welfare	18%
Sociology	Sociology, macro management	4%
Other	Research managers, funders, event organizers	13%

Behavioral Coding. Ideas for what to code were brainstormed while attending a different conference and refined through discussion. Coding themes fell into three broad categories: (1) the frequency with which people approached the microphone and length of time at the microphone, (2) how many distinct remarks were presented and the classification of those remarks as questions or comments, and (3) the tone and demeanor of their time at the microphone. For the full list of coded behaviors, see the codebook (https://osf.io/ze9fh). Two trained research assistants blind to hypotheses watched the conference and coded each behavior with good reliability (0.79 < *Cohen's Kappa* < 0.93). Discrepancies were resolved by a senior researcher.

Participation. Observations were defined as discrete interactions between one question-asker and the speaker during a Q&A session. For each interaction, raters recorded the total time (in seconds) that the attendee and the speaker spoke. Time for attendees was separated by their initial remarks and follow ups. Raters also recorded time for personal introductions. Time recording between the two raters was considered to be discrepant if the differences were above five seconds and were resolved by a third researcher. Time recording differences below five seconds were averaged.

Type of Remark. I categorized remarks given at the microphone into one of four types of contributions: (1) questions, (2) comments, (3) comments with an afterthought of a question at the end (comment-questions), and (4) questions that are subsequently answered in the remark (question-comments; see Table 3 for examples). Types of remarks were coded separately for initial remarks and follow-up remarks. Remarks were coded on the idea-level and not the sentence-level, i.e., an idea spanning across multiple clauses was coded as a single remark. It was possible for attendees to contribute multiple remarks spanning multiple categories, (e.g., “I have a comment and a question.”).

Dominance. Dominance was operationalized along four dimensions. First, dominance was measured by earliness to ask questions (Burgoon & Dunbar, 2006; Galinsky et al., 2003). Earliness was measured as asking one of the first four questions during a Q&A period. This cutoff was created to create a meaningful early chunk because men were overrepresented across ordering.

Next, raters coded for whether the attendee challenged the speaker in their remark. Challenging the speaker was defined as any combination of 1) questioning the validity of a claim made by the speaker, 2) questioning the speaker’s knowledge or expertise, or 3) questioning the integrity or reliability of the presented research. Challenges did not include suggestions for improving research. Rather, this category focused on whether or not the question asker indicated or implied that the speaker was wrong about something.

I also measured the difference between the length of time question-askers spoke from the length of the speaker’s response. I hypothesized this would exemplify a measure of dominance because in a Q&A session, the expert is the speaker and therefore the emphasis and time allotment should be focused on the speaker’s response. To the extent the question-asker speaks longer than the speaker, they are promoting their own status over the speaker. This behavior was measured as the difference between the length of time the speaker spoke from the length of time the question-asker spoke.

Lastly, I defined peacocking as a composite of self-important behaviors: the number of times the attendee talked over the speaker, if the attendee positively referred to their own research, if they challenged the speaker, the amount of time spent introducing themself, not asking a question in their initial remark, and the amount of time spent giving their initial remark (see correlations of all items in Table 2). Each indicator was *Z*-scored and then averaged. Though some of these behaviors occurred infrequently, I rationalized that they could meaningfully be evaluated with similar behaviors. Though reliability was low and unable to be estimated with such a small sample, this may be because several of these behaviors are dichotomous which limits measurement variance and they do not co-occur. That is, when peacocking, individuals may be unlikely to employ multiple actions – instead, engaging in one or a small subset is sufficient to express self-importance. Similarly, if one were to measure enjoyment of drinking alcohol by frequencies of drinking different types of alcohols they would similarly have poor internal reliability due to preferences of different types of alcohols over others. Despite their poor internal reliability, all would contribute to an understanding of the construct of alcohol consumption. It is also possible that peacocking isn’t a coherent construct as we created it.

Table 2. Correlations of peacocking items.

	1.	2.	3.	4.	5.	6.
1. Intro Time	1.00					
2. Initial Question Time	-.26	1.00				
3. No Questions	-.08	.24	1.00			
4. Challenge Speaker	-.17	.60***	.29*	1.00		
5. Positive Self-Reference	.13	.67***	.27	.51***	1.00	
6. Talk over Speaker	-.08	-.11	.01	.09	-.07	1.00

* p < .05, ** p < .01, *** p <.001

Politeness. I categorized the following actions as polite: (1) thanking the speaker for the talk, (2) beginning their remarks with a positive comment about the talk (e.g., "That was an interesting talk."), or (3) thanking the speaker for their response. Question askers were coded as exhibiting politeness if they demonstrated any of these behaviors (0 = Not Polite; 1 = Polite).

Survey. A follow up survey was sent to conference attendees approximately six months after the conference ended. Of the conference attendees, 49% participated (25 men, 19 women, 1 nonbinary person). Questions about comfort in Q&A sessions were included at the end of a survey about experiences at the conference (see OSF for materials: https://osf.io/z4p29/). Survey participants responded to the following 3 questions about the conference they attended and conferences generally: (1) How comfortable were/are you asking questions during Q&A sessions?, (2) How comfortable were/are you sharing comments or opinions during Q&A sessions?, (3) How comfortable were/are you approaching speakers outside Q&A sessions? on 7-point Likert scales (1 = Very Uncomfortable; 7 = Very Comfortable). Participants also answered one question about how often they hold back questions on a 5-point Likert scale (1 = Never; 5 = Always) and were provided the opportunity to describe why they hold back questions if they hold back questions.

Qualitative Coding. Of the survey respondents, 29 participants described why they hold back questions (17 men, 11 women, 1 nonbinary person). Two research assistants blind to the gender of respondents and hypotheses reviewed the responses and pulled out four key themes that occurred with reasonable frequency: anxiety, making space for others to participate, logistical constraints, and preferring one-on-one communication (see Table 3). The two research assistants coded the responses (0 = Absence; 1 = Presence) with good reliability (0.72 < *Cohen's Kappa* < 1), and discrepancies were resolved by a senior researcher.

Table 3. Each construct that was coded, their definitions, and examples of what participants said that fell into each code.

Construct	Definition	Example
Anxiety	Felt insecure or anxious, e.g., being afraid to speak in crowds or worrying about not sounding smart.	*I'm a senior graduate student and often second guess whether my comment is worth asking/saying.*
Make Space	Allowed other attendees to participate over themselves.	*I do not want to dominate the sessions and crowd out others*
Logistics	Had logistical or practical constraints, e.g., being far from the microphone or other aspects of the set up.	*Because time was short/host set limit the number of questions that could be asked during the q/a session.*
One-on-One	Preferred one-on-one conversations over participating in Q&A sessions.	*Mostly not sure whether it is a question I want to announce to the room, I'd rather ask the speaker face-to-face.*

Results

Behavioral Participation. To provide direction on where to focus confirmatory tests based on all of the coded dimensions, for each potential dependent variable, I tested question-asker gender effects, speaker gender effects, and their interactions. Given how small the conference was, I did not analyze behaviors that occurred too infrequently (less than five times across the whole conference), and I determined effects to be promising based on mean differences and trends in the data in addition to significance testing.

Based on the analyses, I determined the following eight analyses of question-asker gender effects to be the most promising hypotheses for confirmatory tests: (1) disproportionate approaches to the mic ($\chi^2(1, N = 50) = 3.53$, $p = .060$, $d = 0.55$, *95% CI* [-0.04, 1.13]), (2) longer questions when at the mic ($b = -0.69$, $Z = -2.74$, $p = .006$, *95% CI* [-1.16, -0.17]), (3) disproportionate total floor time ($\chi^2(1, N = 3374) = 653.49$, $p < .001$, $d = 0.98$, *95% CI* [0.90, 1.06], which was strikingly a 10:1 time ratio when ratio of attendees was 3:1), (4) only say comments at the mic ($b = -0.24$, $t(48) = -1.55$, $p = .128$, *95% CI* [-0.54, 0.06]) which was only done by men, (5) first four to ask questions ($\chi^2(1, N = 22) = 1.57$, $p = .211$, $d = 0.54$, *95% CI* [-0.35, 1.43]) which was limited to the first four to create a meaningful chunk because men are overrepresented across question positions making continuous analyses less informative, (6) difference between amount of time asking question to how much time the speaker responded ($t(12) = 1.71$, $p = .113$, $d = 0.66$, *95% CI* [-0.11, 1.43]), (7) challenging the speaker ($b = -0.04$, $t(48) = -0.29$, $p = .774$, *95% CI* [-0.32, 0.24]) which was almost only done by men, and peacocking ($t(14) = 1.02$, $p = .324$, $d = 0.39$, *95% CI* [-0.37, 1.15]).

Next, I evaluated the components of the peacocking composite. Of the peacocking components, men spent more time asking their initial questions and were less likely to pose a question during their speaking time (see Table 4).

Table 4. Tests of gender differences on each peacocking item.

	Men	Women			95% CI	
	M(SD) or Count	*M(SD)* or Count	Test Statistic	*p*	*Lower*	*Upper*
Full Composite	0.03 (0.56)	-0.13 (0.37)	$t(14) = 1.02$, $d = 0.39$	.324	-0.37	1.15
Intro Time	0.50 (1.40)	1.19 (2.20)	$b = -0.68$, $Z = -0.72$	.469	-2.50	1.15
Initial Question Time	72.17 (48.16)	36.50 (25.27)	$b = -0.68$, $Z = -2.62$	.009	-1.16	-0.14
No Questions	16	0	$b = -0.38$, $t(48) = -2.17$	.035	-0.72	-0.04
Challenge Speaker	7	1	$b = -0.04$, $t(48) = -0.29$	.774	-0.32	0.24
Positive Self-Reference	6	0	$b = -0.14$, $t(48) = -1.13$	.264	-0.39	0.10
Talk over Speaker[1]	1	1				

Frequencies and gender differences in utilization of the different remark types are reported in Table 5. In this sample, men were significantly more likely to provide comments than women. However, there were no significant gender differences in the number of initial remarks ($t(12) = -0.22$, $p = .831$, $d = 0.08$, *95% CI* [-0.67, 0.84]), follow up remarks ($t(8) = -0.94$, $p = .375$, $d = 0.36$, *95% CI* [0.40, 1.12]), or total remarks ($t(9) = -0.95$, $p = .364$, $d = 0.37$, *95% CI* [-0.39, 1.13]).

[1] No test statistics included for this factor because it only occurred twice.

Table 5. Examples of each type of remark, the frequency of each type of remark by gender, and the extent to which the proportion of participation by gender differed from what would be expected by the gender base rates of the conference attendees.

							95% CI	
Category	Example	N_{men}	N_{women}	χ^2 (1)	*p*	*d*	*Lower*	*Upper*
Question	How do you think your research would relate to Social Identity Theory?	29	10	0.17	.676	0.13	-0.50	0.76
Comment	Your work reminds me of Social Identity Theory.	21	1	4.46	.035	0.99	0.03	1.94
Comment - Question	Your work reminds me of Social Identity Theory. Thoughts?	6	0	0.65	.419	0.64	-1.22	2.49
Question-Comment	How do you think your research would relate to Social Identity Theory? Because to me, your second study reminds me of work in that area.	5	0	0.36	.551	0.50	-1.54	2.53
All Remarks		61	12	5.24	.022	0.55	0.07	1.03

Overall, there seemed to be little to no effect of the speaker's gender on any of the potential outcomes. Directionally, question askers were more likely to be polite to speakers who were women compared to men (b = 0.04, $t(47)$ = 0.31, p = .760, *95% CI* [-0.24, 0.32]).

Comfort in Remarking. Next, in a survey we asked conference attendees to report their comfortability participating in several aspects of Q&A sessions. Overall, effects were directionally in our hypothesized direction: men were more comfortable participating in all aspects of Q&A sessions and were less likely to hold back questions. (see Table 6).

Table 6. Results for comfort remarking items.

	Men		Women					95% CI	
	M	*SD*	*M*	*SD*	*t*	*df*	*p*	*Lower*	*Upper*
Comfort Asking Questions at Exploratory Conference	5.28	1.49	4.31	1.65	1.68	24	.106	0.61	-0.12
Comfort Asking Questions in General	5.00	1.78	4.38	1.66	0.99	27	.332	0.36	-0.36
Comfort Providing Comments at Exploratory Conference	5.28	1.45	4.38	1.94	1.40	21	.175	0.51	-0.21
Comfort Providing Comments in General	5.00	1.78	4.23	1.79	1.18	26	.247	0.43	-0.29
Comfort Approaching One-on-One at Exploratory Conference	5.67	1.14	5.15	1.41	1.08	22	.290	0.39	-0.33
Comfort Approaching One-on-One in General	5.17	1.34	4.77	1.64	0.72	23	.480	0.26	-0.46
Hold Back Questions	3.00	0.84	3.31	1.11	0.84	21	.410	0.31	-0.41

Open-Ended Reasons for Holding Back Questions. Of the four themes we identified in the attendee responses, only anxiety showed gender differences in why people hold back questions with most women reporting they hold back questions due to anxiety whereas men displayed half the effect size. Directionally, men were more likely to report wanting to make space for others (and no women indicated this), and women were more likely to report logistics as an inhibiting factor for participation (see Table 7).

Table 7. Results for open-ended reasons for holding back questions.

	Men		Women					*95% CI*	
	M	*SD*	*M*	*SD*	*b*	*Z*	*p*	*Lower*	*Upper*
Anxiety	0.41	0.51	0.82	0.40	1.86	2.01	.044	0.18	3.94
Make Space	0.24	0.44	0.00	0.00	-18.39	-0.01	.995	--	--
Logistics	0.06	0.24	0.27	0.47	1.79	1.45	.146	-0.43	4.89
One-on-One	0.12	0.33	0.09	0.30	-0.29	-0.22	.824	-3.43	2.18

Study 2: Behavioral Participation

Methods

Data Source. Conference 1 was attended by 375 attendees who were 35% women, 63% men, and 2% other or unspecified (see Table 8 for disciplinary representation). There were 193 Q&A interactions across 32 research talks. The sample size was determined by the number of interactions that occurred at the conference. All Q&A interactions were initiated by waiting in line at one of two stationary microphones on either side of the room[2]. Questions asked by moderators were excluded from analyses (4 interactions). Information about attendees was obtained from conference organizers. The Q&A interactions were documented from recordings of the conference posted online. Of the conference speakers, 45% were female and 55% were male.

[2] I was unable to analyze any aspect of gendered behavior in line because the video cameras only captured the speaker on stage.

Table 8. Conference attendees by discipline.

Field	Components	Percent
Biology	Biology, cellular biology	4%
Communications		2%
Data Science	Data science, statistics, information sciences, information technology	6%
Economics	Economics, behavioral economics	3%
Engineering & Math	Bioengineering, Computer science, chemical engineering, computer engineering, electrical engineering, mathematics	12%
Industry	Tech start ups	3%
Law		1%
Medical Sciences	Medicine, epidemiology, radiology, pathology	7%
Metascience & Philosophy	Metascience, meta-research, philosophy, Philosophy of science, science of science	8%
Physics	Physics, astrophysics, applied physics	2%
Political Science		2%
Psychology and Brain Sciences	Psychology, psychiatry, neuroscience, behavioral marketing, micro management, cognitive science, developmental psychology, social psychology, UX	26%
Science Communication	Journalists, freelance writers	3%
Social Sciences	Public health, public policy, health policy	4%
Sociology	Sociology, macro management	3%
Other	Research managers, funders, event organizers	12%
Unidentified		1%

Gender Base Rates of Attendees. To calculate gender base rates of attendees, I used a list of attendees from the conference organizers and coded self-identified gender from several sources: self-identified gender via survey response (38% of attendees), pronouns on personal websites (44% of attendees), pictures or names (17% of attendees), with 2% unidentified.

Question-Asker Gender. Gender presentation of question-askers was coded by two independent raters based on Q&A sessions' recordings and by using vocal cues such as pitch (*Cohen's Kappa* = 1.00). When question-askers introduced themselves, their gender identity was cross-checked with the conference list (18% of question askers). The data were then anonymized.

Qualitative Coding of Conference Sessions. The qualitative coding of the conference Q&A sessions followed the codebook from Study 1. Two independent raters blind to the hypotheses watched all Q&A sessions and indicated whether behaviors occurred in accordance with the codebook. Raters were trained on 10% of the data and then rated the rest of the data. They achieved sufficient interrater reliability across the constructs (*weighted Kappa* for ordinal data = 0.92; *Cohen's Kappa* for categorical data = 0.69). For a full list of recorded behaviors, see Appendix 1.

***Participation*.** Observations were defined as discrete interactions between one question asker and the speaker during a Q&A session. For each interaction, I recorded the total time (in seconds) that the attendee and the speaker spoke. Time for attendees was separated by their initial remarks and follow ups. I also recorded time for personal introductions. Time recording between the two raters was considered discrepant if the differences were above five seconds and were resolved by a senior researcher. Time recording differences below five seconds were averaged.

***Remark Content*.** I categorized remarks given at the microphone into one of four types of contributions: (1) questions, (2) comments, (3) comments with an afterthought of a question at the end (comment-questions), and (4) questions that are subsequently answered in the remark (question-comments). Types of remarks were tracked separately for initial remarks (N = 233) and follow-up remarks (N = 65). Remarks were coded on the idea-level and not the sentence-level, i.e., an idea spanning across multiple clauses was coded as a single remark. It was possible for attendees to contribute multiple remarks spanning multiple categories, (e.g., "I have a comment and a question.").

***Dominance*.** Dominance was operationalized along four dimensions. First, I measured earliness to ask questions (Burgoon & Dunbar, 2006). Earliness was measured as asking one of the first four questions during a Q&A session, meaning that the question asker was either first or second in line to ask a question per microphone. This cutoff was pre-registered.
Next, I coded for whether the attendee challenged the speaker in their remark. Challenging the speaker was defined as any combination of 1) questioning the validity of a claim made by the speaker, 2) questioning the speaker's knowledge or expertise, or 3) questioning the integrity or reliability of the presented research. Challenges did not include suggestions for improving research. Rather, this category focused on whether the question asker indicated or implied that the speaker was wrong about something.

Next, I measured the difference between the length of time question-askers spoke from the length of the speaker's response. Lastly, I created the peacocking composite from Study 1 (see Table 9 for correlations between items). Each indicator was Z-scored and then averaged (α = .44).

Table 9. Correlations of peacocking items.

	1.	2.	3.	4.	5.	
1. Intro Time	1.00					
2. Initial Question Time	.06	1.00				
3. No Questions	-.05	.26***	1.00			
4. Challenge Speaker	-.01	.15*	.35***	1.00		
5. Positive Self-Reference	.13	.32***	.35***	.04	1.00	
6. Talk Over Speaker	-.11	.04	.03	.19**	.01	1.00

* $p < .05$, ** $p < .01$, *** $p < .001$

Politeness. I categorized the following actions as polite: (1) thanking the speaker for the talk, (2) beginning their remarks with a positive comment about the talk (e.g., "That was an interesting talk."), or (3) thanking the speaker for their response. Question-askers were coded as exhibiting politeness if they demonstrated any of these behaviors (0 = not polite; 1 = polite).

Results

Participation Rates. To examine gender differences in participation, I compared rates of participation to the gender base rates of conference attendees. Because men made up 63% of conference attendees, I would expect approximately 63% of questions to be asked by men if there is no gender gap in participation rates. The extent to which men collectively ask more than 63% of questions would suggest disproportionate participation. I used the base rate as defined by the gender of all attendees because this was a single-track conference, meaning all attendees experience the same set of sessions.

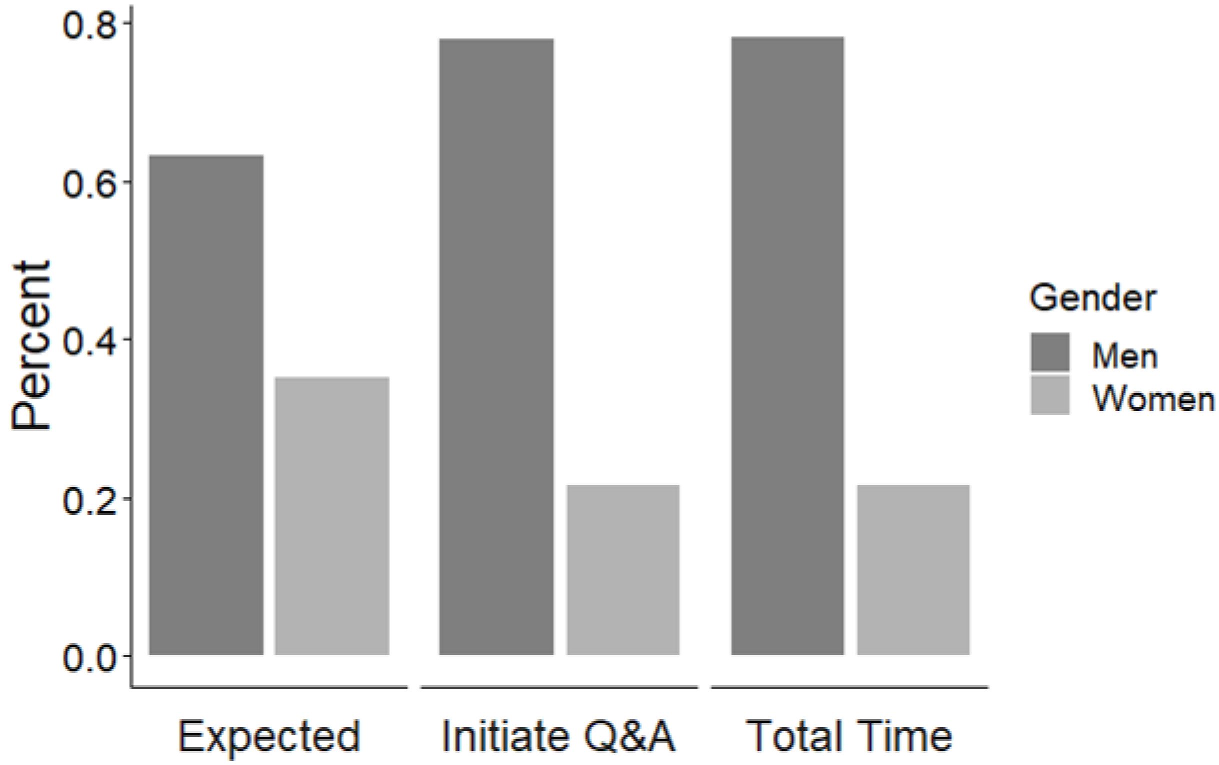
Did men disproportionately initiate Q&A interactions across the conference? Compared to their representation among attendees (63%), men were more likely to approach the microphone and initiate Q&A interactions (78%), $\chi^2(1, N = 189) = 8.41$, $p = .004$, $d = 0.43$, *95% CI* [0.14, 0.72] (see Figure 1).

Were there gender differences in the length of time participants spent talking per Q&A interaction? Per Q&A interaction, men ($M = 66.87$ seconds, $SD = 44.51$ seconds) and women ($M = 66.93$ seconds, $SD = 42.89$ seconds) did not differ in speaking time, $b = 0.00$, $Z = 0.02$, $p = .984$, *95% CI* [-0.20, 0.21].

Did men disproportionately take more of the total Q&A time across the conference? Due to the disparity in volume of questions, men (9897 seconds), compared to women (2744 seconds), consumed a disproportionate amount of the total Q&A time, $\chi^2(1, N = 12{,}641) = 605.64$, $p < .001$, $d = 0.45$, *95% CI* [0.41, 0.48][3]. Base rates determined by conference attendance would predict that men would consume almost twice as much of the Q&A sessions compared to women. In actuality, men consumed more than 3.5 times as much Q&A time.

[3] This analysis includes the total amount of time attendees spent talking at the microphone excluding speaker response time.

Figure 1. Proportion of participation by men and women in Q&A sessions. The first set of columns is the expected rate of participation given the base rates of attendees. The second and third set of columns is the actual participation rate for men and women for initiating Q&A interactions and total time on the floor respectively.



Analysis of Remark Content.

Were types of remarks used disproportionately across the conference by gender? I analyzed the usage of types of remarks across the conference by gender using χ^2 tests comparing the frequency with which men and women contributed remarks to the gender base rates of conference attendees. Men were more likely to ask questions and provide comments while at the microphone in their initial remarks (see Table 10).

Table 10. The frequency of each type of remark by gender, and the extent to which the proportion of participation by gender differed from what would be expected by the gender base rates of the conference attendees. Comment-questions and question-comments occurred too infrequently for meaningful conclusions about gender differences in utilization of these remarks to be drawn from these data.

						95% CI	
Category	N_{men}	N_{women}	χ^2 (1)	*p*	*d*	*Lower*	*Upper*
Question	134	35	8.67	.003	0.46	0.15	0.77
Comment	47	10	3.92	.048	0.54	0.00	1.80
Comment-Question	2	3	0.02	.898	0.10	-1.86	2.07
Question-Comment	2	0	0.00	>.999	0.00	-3.92	3.92
All Remarks	185	48	12.42	<.001	0.47	0.21	0.74

Were there gender differences in the number of distinct points made per Q&A interaction? I analyzed gender differences in the average number of remarks per question-asker using *t*-tests. However, no significant gender differences emerged in the number of initial remarks (Women: M = 1.17, SD = 0.38; Men: M = 1.25, SD = 0.53; $t(88) = 1.07$, $p = .286$, d = 0.19, *95% CI* [-0.16, 0.54]), follow up remarks (Women: M = 0.37, SD = 0.70; Men: M = 0.34, SD = 0.64; $t(60) = 0.23$, $p = .818$, $d = 0.04$, *95% CI* [-0.31, 0.39]), or total remarks (Women: M = 1.54, SD = 0.71; Men: M = 1.59, SD = 0.86; $t(76) = 0.39$, $p = .698$, $d = 0.07$, *95% CI* [-0.28, 0.41]).

Were there gender differences in only saying comments (and no other question type) during initial Q&A interactions per interaction? No significant gender differences emerged in remark types used during initial remarks that were not questions including only providing comments (Women: M = 0.20, Men: M = 0.14; $b = 0.05$, $t(187) = 0.83$, $p = .405$, *95% CI* [-0.07, 0.18]).

Were there gender differences in the omission of questions (providing comments, comment-questions, or question comments but no questions) during initial Q&A interactions per interaction? There were no significant gender differences in failures to ask questions (Women: M = 0.24; Men: M = 0.17; $b = 0.07$, $t(187) = 1.09$, $p = .276$, *95% CI* [-0.06, 0.21]).

Dominance.

Were men disproportionately first to offer remarks in Q&A sessions across the conference? Men (N = 99) were more likely to be among the first four audience members to offer remarks than women (N = 21), $\chi^2(1, N = 120) = 9.27$, $p = .002$, $d = 0.58$, *95% CI* [0.20, 0.95].

Were there gender differences in how challenging or polite participants were to speakers? There were no significant differences by gender in the likelihood of challenging speakers (12% of men, 15% of women), $b = 0.02$, $t(187) = 0.42$, $p = .676$, *95% CI* [-0.09, 0.14]. Nor were there gender differences in politeness of remarks (57% of men, 44% of women), $b = -0.13$, $t(187) = -1.46$, $p = .145$, *95% CI* [-0.30, 0.04].

Were there gender differences in participants out-speaking the speaking during their time at the microphone? There were no significant gender differences on the difference between the length of the remark and the length of the speaker response, $b = 2.12$, $t(159) = 0.22$, $p = .827$, *95% CI* [-16.92, 21.32].

Were there gender differences in peacocking? Utilization of peacocking behaviors was directionally in the opposite direction of what was observed in the exploratory data set, $b = 0.16$, $t(162) = 1.91$, $p = .058$, *95% CI* [-0.01, 0.32]. When evaluating the individual components of the peacocking composite, directionally, women spent more time speaking, but men were more likely to do the other behaviors (see Table 11).

Table 11. Tests of gender differences on each peacocking item.

	Men	Women			*95% CI*	
	M(SD) or Count	*M(SD)* or Count	Test Statistic	*p*	*Lower*	*Upper*
Full Composite	-0.03 (0.49)	0.11 (0.59)	$b = 0.16$, $t(162) = 1.91$	.058	-0.01	0.32
Intro Time	1.85 (2.19)	2.35 (2.18)	$b = -0.41$, $Z = -1.06$	.290	-1.17	0.35
Initial Question Time	60.19 (42.23)	60.21 (42.07)	$b = 0.00$, $Z = 0.01$	.992	-0.21	0.22
No Questions	25	10	$b = 0.07$, $t(187) = 1.09$	.276	-0.06	0.21
Challenge Speaker	18	6	$b = 0.02$, $t(187) = 0.42$	.676	-0.09	0.14
Positive Self-Reference	3	2	$b = 0.03$, $t(187) = 1.00$	.317	-0.03	0.08
Talk over Speaker	23	9	$b = 0.06$, $t(187) = 0.97$	.335	-0.07	0.19

Exploratory Analysis of Speaker Gender Effects. I tested if speaker gender predicted attendee Q&A behaviors. Of the behaviors described above, only one exhibited significant speaker gender effects and had directional consistency across the exploratory sample (Study 1) and the confirmatory sample (this study): attendees were more likely to engage in polite behaviors when speakers were women (66%) than men (42%), $b = 0.22$, $t(196) = 3.20$, $p = .002$, *95% CI* [0.09, 0.36].

Interim Discussion

Study 2 provided behavioral evidence that men are more likely to participate during Q&A sessions than women relative to their attendance rate at the conference. This difference extended to both asking questions and providing comments. While I observed a tendency for men to be among the first four participants in Q&A, a sign of dominance, I did not find that men were more challenging or less polite in their remarks relative to women. This study builds upon past research by testing with greater specificity gender differences in *how* Q&A session time is used.

Next, in Study 3, I examined how these behavioral differences relate to men's and women's psychological experiences during Q&A sessions. Specifically, I investigated self-reported barriers to speaking by asking conference attendees to rate their comfort in participating in asking questions, commenting, and approaching speakers after Q&A sessions, as well as their tendencies to hold back questions and to fear backlash for participating. Questions were asked of both the focal conference and conferences generally to ensure observed effects were not due to an idiosyncrasy with the focal conference. Next, I measured attendees' qualitative responses describing why they hold back questions. I then tested the replicability of both the quantitative and qualitative findings using openly available data. Incorporating both quantitative and qualitative components provides insight from the perspective of conference attendees. At the focal conference in Study 3, women were the numerical majority, meaning any effects found are not simply driven by the fact that men are the numerical majority, as in Study 2.

Study 3: Self-Report

Method

Participants. Recruitment emails were sent to a psychology society's listserv approximately six months after the annual conference. Conference 2 was for a sub-discipline of psychology in which women are the numerical majority of conference attendees (61% women). It is a large, multi-track conference spanning several days, and serves as the main conference for researchers within this sub-discipline. Participants can choose between many symposium sessions, poster sessions, and professional development workshops scheduled throughout the conference.

Conference attendees were offered the chance to win $35 via a drawing for their participation. Of the 4208 conference attendees, 284 attendees participated in the survey and 234 attendees completed the focal items for this study. Sample size was determined by the maximum number of conference attendees we could recruit within two weeks. Participants were excluded from analyses if they did not fall within the gender binary or did not provide gender information. Survey participants were 69% women, 28% men, and 1% nonbinary, and 2% other or did not disclose, and were predominantly White (59%) and 13% Asian, 9% Latinx, 8% Multiracial, 3% Middle Eastern/North African, 2% Black, <1% other, and 5% did not disclose. Additionally, survey participants were predominantly graduate students (51%) and 3% undergraduates, 9% postdocs, 32% professors, 3% in industry, and 2% other. Analyses comparing the demographics of our sample to that of the conference attendees suggest our sample is not significantly different from the full sample in regards to race ($\chi^2(9, N = 232) = 15.03$, $p = .090$, $d = 0.53$, *95% CI* [0.26, 0.79]) or gender ($\chi^2(5, N = 233) = 7.77$, $p = .170$, $d = 0.37$, *95% CI* [0.11, 0.63]), but included more early career scientists and fewer full professors ($\chi^2(3, N = 234) = 73.34$, $p < .001$, $d = 1.35$, *95% CI* [1.04, 1.66]).

Materials. The items used in this investigation were added to a survey for a separate investigation on broader conference participation (see https://osf.io/a84hg for preregistration and https://osf.io/8p3vm for full materials).

Comfort in Remarking. Survey participants responded to three questions about how comfortable they felt engaging in Q&A sessions at Conference 2: (1) asking questions, (2) sharing comments or opinions, (3) approaching speakers outside Q&A sessions on 7-point Likert scales (1 = Very Uncomfortable; 7 = Very Comfortable). I also asked the same three questions about participating in Q&A sessions at conferences generally. As the items were highly correlated ($rs_{specfic} > .57$; $rs_{general} > .58$), they were combined into a comfort index for ease of reporting ($\alpha_{specific} = .85$, $\alpha_{general} = .85$).

Fear of Backlash. Participants indicated how afraid they were of receiving backlash for participating both at Conference 2 and conferences generally on 5-point Likert scales (1 = Never; 5 = Always). I did not define backlash because I assumed it is commonly understood to be fear of negative appraisals or retribution by others.

Propensity to Hold Back Questions. Participants indicated how often they hold back questions on 5-point Likert scales (1 = Never; 5 = Always).

Qualitative Coding.

Participants and Procedure. Of the survey participants, 198 participants (138 women, 60 men) responded to the following open-ended question: If you hold back questions during Q&A sessions, why? On average, participants wrote 19-word responses, which referred to, on average, approximately 1.1 of the 4 reasons in the codebook (described below).

Codebook. Coding themes were identified by two research assistants blind to hypotheses and participant gender by induction using the exploratory conference. The coding themes were then applied to the data from Conference 2. The four themes that arose most frequently included: anxiety, making space for others, logistics, and preferring one-on-one communication (see Table 4)[4]. Two independent raters blind to hypotheses and participant gender coded the open responses according to the codebook. Each code was given either a 0 for absence or 1 for presence, and discrepancies were resolved by a senior research member. Independent raters achieved high inter-rater reliability for both data sets ($Cohen's\ Kappa_{main}$ = 0.77; $Cohen's\ Kappa_{replication}$ = 0.79). Additional exploratory measures were coded and included in the codebook (see Appendix 2).

Conceptual Replication. Lastly, I attempted to replicate the quantitative and qualitative results by applying the codebook to a publicly available data set (Carter et al., 2018), which surveyed academics about their participation in departmental seminars. Specifically, participants were asked to rate the following factors related to discomfort that might prevent them from asking questions on a 5-point Likert scale (1 = Not at all Important; 5 = Extremely Important): (1) Worried that I had misunderstood the content; (2) Couldn't work up the nerve; (3) Not sure whether the question was appropriate; (4) The speaker was too eminent/intimidating; (5) Worried that I was not clever enough to ask a good question[5]. They were also given the opportunity to write in factors that were not already listed. A relatively large sample provided Likert responses (N = 466, 277 women, 189 men) and a smaller subset provided additional factors (N = 104, 61 women, 42 men).

To replicate the quantitative analyses, I combined the five items to create a single discomfort composite (α = .84). I then coded the additional factors described in open-ended responses according to our codebook to replicate our analyses with the open-ended responses.

Results

Comfort in Remarking.

Were there gender differences in comfort participating? At the focal conference, women reported less comfort participating in Q&A sessions compared to men (Women: M = 3.77, SD = 1.51; Men: M = 4.76, SD = 1.38; $t(124)$ = 4.73, p < .001, d = 0.70, *95% CI* [0.40, 1.00]). Women also reported less comfort participating in Q&A sessions than men in conferences generally (Women: M = 3.82, SD = 1.49; Men: M = 4.69, SD = 1.39; $t(124)$ = 4.17, p < .001, d = 0.62, *95% CI* [0.32, 0.91]; see Table 12 and Figure 2 for individual items).

---

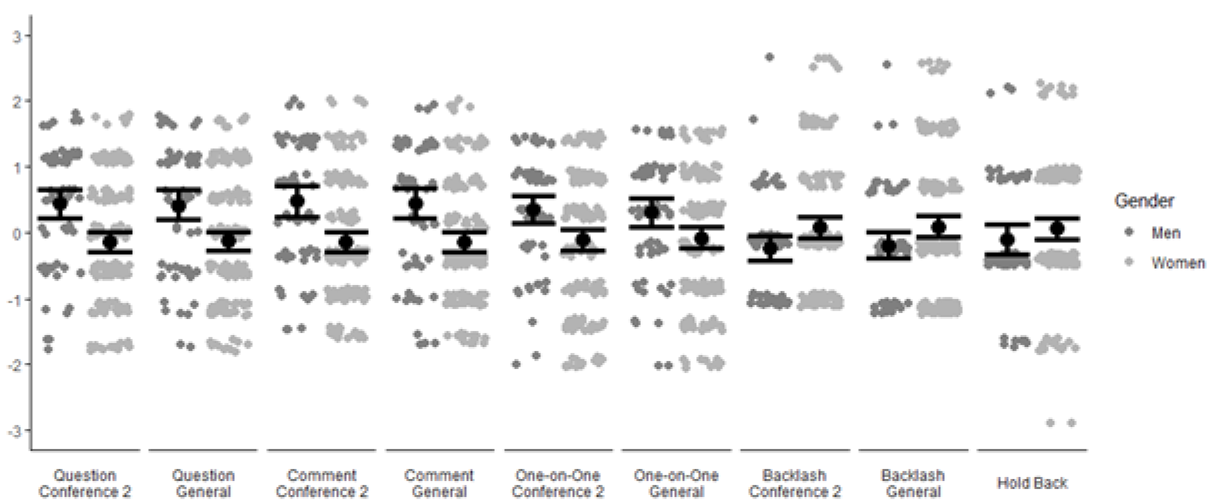
[4] Note, these are spontaneous attributions and reflect what first comes to mind. Thus, responses may vary from when asked about particular factors via Likert-type questions.

[5] Three additional items were collected but not analyzed in this investigation because they were logistical and were unrelated to discomfort: (1) Not enough time, (2) Not my field; (3) I was meeting the speaker later/asked after the talk had ended.

Table 12. Results for individual items in the comfort remarking composite.

	Men		Women		*t(df)*	*p*	*d*	*95%CI*	
	M	*SD*	*M*	*SD*				*Lower*	*Upper*
Comfort Asking Questions at Conference 2	4.74	1.62	3.73	1.71	4.11 (117)	<.001	0.62	0.32	0.91
Comfort Asking Questions in General	4.72	1.65	3.76	1.70	3.89 (120)	<.001	0.57	0.28	0.87
Comfort Providing Comments at Conference 2	4.42	1.67	3.34	1.62	4.33 (108)	<.001	0.65	0.35	0.95
Comfort Providing Comments in General	4.44	1.67	3.46	1.62	4.01 (113)	<.001	0.59	0.30	0.89
Comfort Approaching One-on-One at Conference 2	5.10	1.51	4.25	1.85	3.49 (136)	<.001	0.52	0.22	0.82
Comfort Approaching One-on-One in General	4.92	1.55	4.23	1.77	2.89 (132)	.004	0.43	0.14	0.72

Figure 2. Self-reported comfort participating in Q&A by gender. Data were standardized due to differing scale lengths. Bars are 95% confidence intervals.



Were there gender differences in fears of backlash? Women were also more likely to fear backlash both at the focal conference (Women: M = 2.21, SD = 1.16; Men: M = 1.86, SD = 0.86; $t(153) = 2.50$, p = .013, d = 0.37, *95% CI* [0.08, 0.67]), and at conferences in general (Women: M = 2.33, SD = 1.15; Men: M = 2.00, SD = 0.90; $t(150) = 2.25$, p = .026, d = 0.33, *95% CI* [0.04, 0.62]).

Were there gender differences in holding back questions? Despite reporting less comfort and greater fears of backlash, women did not report holding back questions more than men (Women: M = 3.35, SD = 0.81; Men: M = 3.22, SD = 0.70; $t(137) = 1.23$, p = .219, d = 0.18, *95% CI* [-0.11, 0.47]). Given the stark gender differences in Q&A session behavior observed in Study 1, it is surprising that women did not report holding back questions more than men. We turned next to analyzing whether gender differences emerge in self-reported reasons for *why* conference attendees hold back questions when they do.

Open-Ended Reasons for Holding Back Questions. Using logistic regressions, we analyzed gender differences in the presence of spontaneous attributions of four reasons for holding back questions: anxiety, making space for others, logistics, and preferring one-on-one communication.

Were there gender differences in presence of discomfort? Women were more likely than men to indicate that anxiety impacts their Q&A behavior (Men: M = 0.50, Women: M = 0.66, b = 0.65, Z = 2.06, p = .039, *95% CI* [0.03, 1.27]).

Were there gender differences in indicating trying to make space for others? Men as compared to women were more likely to report holding back questions to make space for others to do so (Men: M = 0.30, Women: M = 0.12, b = -1.18, Z = -3.04, p = .002, *95% CI* [-1.95, -0.42]).

Were there gender differences in the experience of logistics as a barrier to participate? Logistic barriers did not prove to be a greater barrier for men or women (Men: M = 0.22, Women: M = 0.17, b = -0.32, Z = -0.81, p = .416, *95% CI* [-1.07, 0.47]).

Were there gender differences in preferring to ask questions one-on-one? In contrast to the Likert self-reports indicating women are less comfortable approaching the speaker one-on-one, we did not find gender differences in spontaneous mentions of preferring to ask questions in one-on-one conversations after talks (Men: M = 0.18, M_{women} = 0.13, b = -0.39, Z = -0.94, p = .346, *95% CI* [-1.20, 0.45]).

Conceptual Replication.

Were there gender differences in comfort participating? Using the discomfort composite, women reported discomfort impacted why they hold back questions more than men (Women: M = 2.99, SD = 0.94; Men: M = 2.52, SD = 0.94; $t(402) = 5.35$, $p < .001$, d = 0.48, *95% CI* [0.30, 0.66]).

Were there gender differences in the open-ended reasons? In the coding of open-ended responses, only the "make space for others" effect replicated. Men were more likely than women to say that they hold back questions to make space for others to participate (Men: M = 0.29, Women: M = 0.11, b = -1.13, Z = -2.14, p = .033, *95% CI* [-2.21, -0.11]). No significant gender differences emerged in anxiety (Men: M = 0.45, Women: M = 0.38, b = -0.31, Z = -0.76, p = .445, *95% CI* [-1.11, 0.49]), logistics (Men: M = 0.07, Women: M = 0.18, b = 1.05, Z = 1.53, p = .125, *95% CI* [-0.19, 2.59]), or preferring one-on-one conversations (Men: M = 0.00, Women: M = 0.03, b = 17.18, Z = 0.01, p=.995). The absence of a difference on the discomfort measure could be because these responses were meant to supplement factors already present in the question which already included similar Likert-scale items.

Discussion

Academic conferences are an important forum for sharing new research. Q&A sessions are an important part of that forum, providing an opportunity for audience members to comment on presented findings. To the extent men engage more than women, men continue to have more influence over the direction of science. After exploratory analyses in Study 1, in two studies, I found that men as audience members consume a disproportionate amount of conversational space in Q&A sessions and some evidence as to why this effect occurs. In Study 2, I replicated the effect that men are more likely to ask questions and found that they are more likely to be among the first few to do so. They did not, however, spend significantly more time talking or to provide more remarks when at the microphone compared to women. The effects of men's dominance in Q&A sessions seem to be driven by their greater willingness to jump into the discussion rather than in how they communicate while at the microphone.

In Study 3, women reported feeling less comfortable participating in all aspects of Q&A sessions, whether that be for conferences generally or the specific conference they attended in which women were in the numeric majority. Women also reported being more likely to fear backlash because of their participation. This finding extends the work of Brescoll (2011) by showing gender differences even among junior scholars for whom, presumably, egalitarian beliefs are especially strong. Yet, men and women did not significantly differ in the propensity to hold back questions, though they differed in reasons why they hold back questions. Women reported being more likely to hold back questions because they were anxious whereas men reported being more likely to hold back questions to make space for other people to participate. This result suggests that if men were not regulating their behavior, they would dominate even more. Though some credit can be given to men for considering others, it also raises the question of how much bigger the gender gap in Q&A participation might have been in the past.

In addition to the key analyses, I found that both men and women were more likely to be polite to female speakers. This could be a case of benevolent sexism (Dardenne et al., 2007; Glick & Fiske, 1996; Hopkins-Doyle et al., 2019). Alternatively, people may be choosing to match their behavior to stereotypically feminine ways of communication to appease women's apparent anxiety (Bowles & Flynn, 2010; Burgoon et al., 1983; Kray & Thompson, 2004). Future research is necessary to better understand why question-askers are more polite to women speakers.

Possible Sources of Participation Disparities

The findings are consistent with gender disparities in participating in academic spaces, such as in classrooms (Aguillon et al., 2020; Lee & Mccabe, 2021) and conferences in other STEM fields (Hinsley et al., 2017; Käfer et al., 2018; Pritchard et al., 2014; Telis et al., 2019). These data provide some possible explanations for the disparities. For instance, the survey data suggest that women, compared to men, are less comfortable asking questions and providing comments in Q&A sessions. Furthermore, women fear backlash more than men for their contributions. Overall, these concerns could decrease women's participation in Q&A sessions.

Interestingly, while men and women reported different reasons for holding questions back during Q&A sessions, they did not differ in their likelihood of holding back questions. It could be that men genuinely have more questions than women after watching a research talk. I have no data to suggest why that might be. It could also be that, relative to women, men come up with questions more quickly, set lower bars for what they deem to be worth asking, or that due to feelings of stereotype threat, women have less cognitive bandwidth to generate questions in the moment (Schmader, 2010). The results from Study 2 are consistent with each of these explanations, but do not allow us to distinguish between them.

Constraints on Generality

A limitation of this work is that the behaviors from the live conference were from a conference predominantly attended by men. It could be that gender gaps in participation rates are exacerbated when women are in the numeric minority (Kanter, 1977). However, in Study 3 women were in the numeric majority but still reported feeling less comfortable participating in Q&A sessions compared to men. Additionally, past research has found women were underrepresented in question askers in genetics and biogenetics regardless of whether women were in the numeric minority or numeric minority (Telis et al., 2019). It does not seem plausible that gender base rates of attendees at the conference can completely explain our observed effects. That said, due to the overrepresentation of men in higher professor ranks, there could be a conflation between status due to gender and status due to rank. Future research should compare participation rates between conferences that vary in the gender representation at higher professorial ranks. While the present research focused on gender, in the future, it will be important to study how gender intersects with race and other minoritized identities in shaping participation rates.

This work is also limited by the type of Q&A sessions examined. The conferences did not include moderated Q&A, which could impact participation rates due to how fair conference attendees perceive the moderator to be with equitable calling behavior. Additionally, needing to walk to a microphone could pose a barrier, and women's participation rates could increase if they only needed to raise their hands (Chapman et al., 2016; Leventhal et al., 1965). Finally, I did not include behavioral analysis of a large, multi-track conference, where attendees may feel a greater cover of anonymity that increases the gender diversity of who participates in Q&A. Conversely, women may feel more anxiety in larger conferences due to the larger audiences. Further, session-by-session variation and attendance may impact participation rates.

It is also worth noting that the participation rate for Study 3's survey was particularly low. The recruitment materials described the survey as about experiences of the conference as a whole and how it impacted their work and connections with other researchers. It seems likely that those most interested in conferences would be most likely to participate, however it is also possible that women who were more anxious selectively completed the survey. The findings, however, converge with prior research on gender differences in participation in Q&A sessions in departmental seminars using similar methods (Carter et al., 2018), which reduces the concern that Study 3's results reflect a selection bias.

Q&A sessions provide an opportunity for researchers at all levels to influence science by providing researchers with feedback. In this research, I found, relative to women, men were more likely to participate in Q&A sessions but not in how they participated while at the microphone. Understanding the impacts of disproportionate participation by men in these spaces and adjusting structures to encourage women's participation could increase the diversity of voices in our dialogues about science.

Chapter 2: Fears of Backlash

Introduction

Across fields, question and answer (Q&A) sessions at conferences are dominated by men asking questions – over and above what would be expected based on attendance (Hinsley et al., 2017; Käfer et al., 2018; Pritchard et al., 2014; Schmidt & Davenport, 2017; Telis et al., 2019).

Gender gaps in participation are driven by men being more likely to start Q&A interactions rather than the length of time or what they say at the mic (Study 2). Women report having more anxiety and lower self-confidence when it comes to participating in Q&A sessions (Carter et al., 2018; Study 3). Less explored, however, is the particular role fears of backlash from others plays in Q&A participation. Fears of backlash are most likely to occur when the possibility to act in counter-stereotypic ways emerges, like when women are being assertive and agentic (Eagly & Karau, 2002; Heilman et al., 2004; Heilman, 2012; Rudman & Glick, 1999). To date, there is a large body of research demonstrating backlash for women acting in counter-stereotypic ways, but much less work on the corresponding fears women have for acting in ways that could open a potential for backlash. This chapter tests fears of backlash, uniquely held by women, as a mechanism driving gender participation in Q&A sessions.

Gender Stereotyping and Backlash

Categorization and stereotyping are fundamental aspects of the human experience – it functions to help people process complex information but often leads to oversimplifications (Taylor et al., 1978). In the United States, the origins of stereotypes of women and men are linked to traditional gender social roles of women as caretakers and men as providers (Wood & Eagly, 2012). Women are stereotyped as being higher in warmth and communal tendencies and lower in competence and agentic tendencies than men (Eagly & Steffen, 1984; Fiske et al., 2002). The stereotypical qualities seen as necessary for successful workers and leaders (i.e., agentic and masculine) are not congruent with the stereotypes for women (Heilman, 2012; Koenig et al., 2011).

Past research shows that women experience backlash for displaying agentic qualities (e.g., advocating for themselves, asking for things, taking up space, etc.) because they run counter to what would be expected by gender stereotypes of women (Eagly & Karau, 2002; Heilman et al., 2004; Heilman, 2012; Rudman & Glick, 1999). For example, women who are assertive in negotiations experience social backlash and are rated as less likable than men (Amanatullah & Tinsley, 2013). Additionally, women who are assertive are viewed more negatively for initiating salary negotiations (Bowles et al., 2007). During the hiring process and as managers, agentic women are perceived as competent but receive backlash in perceptions of their social skills and are seen as more hostile (Heilman & Okimoto, 2007; Rudman, 1998). These perceptions are then used to differentially deny jobs to agentic women (Phelan et al., 2008). Female professors are rated more negatively from high-status departments and are perceived to be tougher graders compared to their female colleagues at low-status departments (Fisher et al., 2019). The impacts of backlash even begin early. Gender stereotype-violating children as young as preschoolers faced social backlash from adults compared to their stereotype-congruent peers (Sullivan et al., 2018). The costs of backlash on women are far reaching and serve to keep women in lower status positions (Fiske, 1993; Ridgeway & Correll, 2004; Rudman et al., 2012).

Fears of Backlash

Correspondingly, women are aware of the potential of backlash for taking up space and modulate their behavior to reduce the degree to which they subject themselves to backlash. According to the Backlash-Avoidance Model, women are more likely to fear backlash for self-promoting activities because they are gender stereotype-incongruent and will then be less active in pursuing these goals and act with greater caution (Moss-Racusin & Rudman, 2010). Women who have the potential to be seen as deviants fear backlash and alter their behavior to be more stereotype-congruent (Rudman & Fairchild, 2004). As a result, self-promoting women are then less successful in achieving their self-promotional aims because they less-actively advocate for themselves and wield the power of their positions (Moss-Racusin & Rudman, 2010). Eliciting fears of backlash functions as a gender hierarchy maintenance tool (Brescoll et al., 2018; Rudman & Fairchild, 2004).

While there has been substantial research on how backlash is enacted, less is known about how the *anticipation* of backlash impacts behavior. Within the realm of negotiations, women reduce their assertiveness when they negotiate for themselves (as opposed to others) and are less likely to initiate salary negotiations due to nervousness and to reduce the opportunities for retaliation (Amanatullah & Morris, 2010; Bowles et al., 2007). Women are less likely to opt-in to be considered for promotions as this is seen as self-promotional and immodest (He et al., 2019). Lastly, women limit the extent to which they delegate tasks to others in fear of backlash, thereby reducing the extent to which they exact their power over others (Akinola et al., 2018). The elicitation of these fears due to structures and individuals as a result of witnessing and experiencing instances of backlash is an additional mechanism to maintain the gender hierarchy status quo (Phelan & Rudman, 2010).

Participation in Q&A sessions aligns with the Backlash-Avoidance model. Within the context of academia, knowledge is the currency by which scholars are judged and valued (Walker et al., 2010), and question-asking can be used as an opportunity to display the knowledge one has or to be judged by others on the merits of asking the question (Cialdini, 2001). By opting-in to participate in Q&A sessions, question-askers simultaneously set themselves up for self-promotion and the potential for group-level scrutiny. Due to the historic inaccessibility of higher education to women, academia continues to be a male-dominant space in which men are more likely to feel like they belong (Leslie et al., 2015). Women, thus, face stereotype incongruency across two fronts, both in self-promotion and in the setting itself. It would then follow according to the theory that women would be less likely to participate in this forum. Consistent with this theory, women do self-report higher fears of backlash compared to men in Q&A settings (Study 3).

Present Studies

Past work has tested links between fears of backlash and inhibited behavior. This investigation expands the understanding of the link between fears of backlash and behavior by manipulating fears of backlash and testing the influence of several factors that could be the source of the fears. Across four studies, I test the explanatory role of fears of backlash in gendered participation in Q&A sessions. I hypothesize that women will be more likely to have fears of backlash and that these fears translate to lowered participation rates. By altering aspects of the social context, I hypothesize that gender differences in participation would be attenuated.

In Study 4, I test whether fears of backlash have explanatory value in gender differences in question-asking intentions using a hypothetical scenario. Next, I examine in Study 5 the influence of participants' perceptions of the audience by testing the impact of a welcoming community on question-asking intentions and fears of backlash. I then build on Study 5 in Study 6 by testing not only the impact of the audience but the nature of what would be said in the question. In Study 7, with a conference simulation, I test the impact of how public the question-asking is on question-asking behavior and related fears of backlash.

Study 4: Correlational Test of Fears of Backlash on Question-Asking Intentions

Study 4 tests the relationship between question asking intentions and fears of backlash in a controlled environment. This study builds upon the self-report data in Study 3 by standardizing the situation participants consider and testing if the behaviorally-observed gender differences can be replicated using hypothetical scenarios. I hypothesize that (1) in a controlled experimental hypothetical setting, men will report a greater likelihood to participate compared to women, and (2) fears of backlash correlated with behavioral intentions for women. This was tested in two preregistered samples (Study 4A: https://aspredicted.org/qp5jf.pdf, Study 4B: https://aspredicted.org/2yz2b.pdf).

Study 4A

Method

Participants. Participants included 191 workers (88 men and 103 women) from Amazon's Mechanical Turk. On average, participants were 26 years old (*SD* = 3 years). Some participants were excluded for pre-registered criteria: 4 participants for being outside of the gender binary, and 6 participants for having never enrolled in college.

Procedure. Participants first answered what their field of study is to target the rest of the study towards their interests. They then read a description about a department's colloquium series with text tailored to their field of study:

> *Every year, [piped text] department invites speakers to come and present about their work.*
>
> *For example, last year [piped text department] invited [speaker from last year] to talk about [x] and [speaker from last year] to talk about [y].*
>
> *Typically, the speaker presents for 45 minutes leaving about 15 minutes for a question-and-answer period. During the question-and-answer period, attendees either raise their hands to be called on or approach a central microphone and wait in line. These talks are open to everyone and tend to include, on average, 100 attendees varying from undergraduates to full professors.*
>
> *Imagine you see a notice for a guest speaker that is particularly interesting to you, so you decide to attend. The talk made you think and left you with some questions.*

Question-Asking Intentions. Participants answered questions about their likelihood to participate in the Q&A session on 7-point Likert scales (1 = Not at all; 7 = Extremely): (1) How likely would you be to ask your question? (2) How determined would you be to ask your question? (3) How eager would you be to share your thoughts? The three items were highly correlated (*r*s > .74) and were averaged into one index for question-asking intentions (a = .92).

Anticipated Backlash Boundaries. Participants answered questions about where they think the threshold for backlash would be for their participation in the Q&A session: (1) How many questions do you think you could reasonably ask without being considered too demanding? (2) How much time do you think you can reasonably take without being seen as too pushy? Both of these items are count data and contain outliers to various degrees. Responses greater than 2.5 SD above the mean or less than 2.5 SD below the mean were excluded from analyses on these measures throughout this chapter.

Fear of Backlash Scale. Participants completed a measure of their fears of backlash (Rudman & Fairchild, 2004) through their agreement to the following six items on 7-point Likert scales (1 = Not at all likely, 7 = Extremely likely; a = .93): (1) Others might think I'm odd, (2) I might be disliked, (3) Others might think I'm too confident, (4) Others might think I'm too assertive, (5) I might be called vain, (6) Others might think I'm too self-promoting.

Results. When considering whether or not to ask a question, men (M_{men} = 4.40, SD_{men} = 1.55) had greater question-asking intentions than women (M_{women} = 3.89, SD_{women} = 1.58; $t(185)$ = 2.27, p = .025, d = 0.33, *95% CI* [0.04, 0.62]). However, there were no significant differences in fears of backlash (M_{men} = 3.13, SD_{men} = 1.44; M_{women} = 3.42, SD_{women} = 1.46; $t(185)$ = 1.36, p =

.176, d = 0.20, *95% CI* [-0.09, 0.48]), the anticipated number of questions that could be asked before experiencing backlash (M_{men} = 3.06, SD_{men} = 1.37; M_{women} = 3.41, SD_{women} = 1.48; t(173) = 1.13, p = .260, d = 0.17, *95% CI* [-0.12, 0.46]), or the anticipated amount of time that could be taken before experiencing backlash (M_{men} = 71.82, SD_{men} = 58.16; M_{women} = 71.43, SD_{women} = 47.65; b = -0.01, Z = -0.05, p = .96, *95% CI* [-0.22, 0.21]).

Correlations between all of the measures were checked with the whole sample and separated by gender (see Table 13). Question asking intentions were related to anticipating more questions could be asked before experiencing backlash for all participants (r(179) = .30, p < .001).

Table 13. Correlations between the key collected dependent variable items. The top halves of the tables are correlations from Study 4A and the bottom halves of the tables are from Study 4B.

	All				Men				Women			
	1.	2.	3.	4.	1.	2.	3.	4.	1.	2.	3.	4.
1. Question-Asking Intentions		-.12	.30***	.12		-.04	.36***	.20		-.15	.21*	.04
2. Fear of Backlash	-.18***		.07	-.02	-.16*		.14	<.01	-.13		.03	-.04
3. Number of Questions	.12*	<.01		.13	.15*	.07		.10	.10	-.09		.17
4. Number of Seconds	.05	-.08	.23***		.05	-.04	.24***		.05	-.11	.22***	

Study 4B

To ensure that the null effects observed in Study 4A were not due to being underpowered, I collected a larger sample size and replicated it in Study 4B. Sample size was determined by being 2.5x greater than the sample size of Study 4A (Simonsohn, 2015).

Method.

Participants. Participants included 489 workers (245 men and 244 women) from Prolific. On average, participants were 26 years old (*SD* = 8 years). Some participants were excluded for pre-registered criteria: 4 participants for being outside of the gender binary, and 9 participants for having never enrolled in college.

Procedure. This study was an exact replication of Study 4A. The three question-asking intention items were highly correlated (rs > .74) and were averaged into one index for question-asking intentions (a = .92). Five items from the Rudman & Fairchild (2004) fear of backlash scale were accidentally dropped in Study 4A. Items were also altered to be questions instead of statements: (1) Would you be afraid that others would think you were odd? (2) Would you be afraid that you might be disliked? (3) Would you worry that others might think you are too confident? (4) Would you worry that others might think you are too assertive? (5) Would you worry about being called vain? (6) Would you worry that someone of your gender should not be self-promoting? (7) Would you worry about being labeled negatively? (8) Would others be likely to (negatively) tease you? (9) Would you worry that someone might think you are acting out of character for someone of your gender? (10) Would you be concerned that you won't be liked because you have acted out of character for someone of your gender? (11) Would you feel proud? The item "Would you feel proud?" did not correlate ($|r|$s < .07) with any of the items in the scale and created errors when trying to run reliability analyses. This item was dropped from the scale for the study and all remaining studies. There was good reliability using the remaining 10 items (a = .92).

Results. When considering whether or not to ask a question, men (M_{men} = 4.39, SD_{men} = 1.63) had greater question-asking intentions than women (M_{women} = 3.93, SD_{women} = 1.61; $t(487) = 3.13$, $p = .002$, $d = 0.28$, *95% CI* [0.10, 0.46]). Women had greater fears of backlash related to asking question compared to men (M_{men} = 2.13, SD_{men} = 0.97; M_{women} = 2.70, SD_{women} = 1.09; $t(480) = 6.19$, $p < .001$, $d = 0.56$, *95% CI* [0.38, 0.74]). There were no significant gender differences on the number of questions that could be asked before experiencing backlash (M_{men} = 1.88, SD_{men} = 0.58; M_{women} = 1.94, SD_{women} = 0.55; $t(478) = 1.08$, $p = .279$, $d = 0.10$, *95% CI* [-0.08, 0.28])[6], or on the anticipated amount of time that could be used before experiencing backlash (M_{men} = 64.08, SD_{men} = 48.94; M_{women} = 62.38, SD_{women} = 47.94; $b = -0.03$, $Z = -0.39$, $p = .698$, *95% CI* [-0.16, 0.11]).

Correlations were checked between all of the measures with the whole sample and separated by gender (see Table 13). Question asking intentions was related to anticipating more questions could be asked before experiencing backlash for all participants ($r(460) = .12$, $p = .010$), and negatively related to fears of backlash ($r(460) = -.18$, $p < .001$).

Last, I tested a pathway that fears of backlash would explain the variance between gender (men = -0.5, women = 0.5) and question-asking intentions. Pathways were modeled using mediation models with 10,000 bootstrapped samples with the LAVAAN package in R (Rosseel, 2012). The predicted indirect path through fear of backlash was significant, $b = -0.14$, $SE = 0.05$, $Z = -2.96$, $p = .003$, *95% CI* [-0.23, -0.05], reducing the magnitude of the direct effect, $b = -0.32$, $SE = 0.15$, $Z = -2.14$, $p = .032$, *95% CI* [-0.62, -0.03] (see Table 14).

Table 14. Path analysis estimates for the effect of question-asking intentions via fears of backlash.

Effect	Estimate	SE	Z	p	95% CI
Gender → Fear of Backlash (a)	0.58	0.09	6.20	<.001	0.39, 0.76
Fear of Backlash → Question Asking Intentions (b)	-0.24	0.07	-3.37	.001	-0.38, -0.10
Total Effect (c)	-0.46	0.15	-3.13	.002	-0.75, -0.17
Direct Effect (c')	-0.32	0.15	-2.14	.032	-0.62, -0.03
Indirect Effect	-0.14	0.05	-2.96	.003	-0.23, -0.05

Interim Discussion

Across the two samples, women reported lower question-asking intentions compared to men. In Study 4B, women also reported greater fears of backlash, and these fears explained the

[6] The negative binomial model to handle the count data could not converge, so a *t*-test was used instead.

relationship between gender and question-asking intentions. Of note, the pathway analysis is cross-sectional and correlational, and causality cannot be ascribed to the path. The null effects in Study 4A were likely due to being underpowered to detect the effects. For the remainder of the chapter, I attempt to manipulate fears of backlash to demonstrate its causal impact on question-asking behavior.

Study 5: Impact of Audience

In Study 5, I test the impact of perceptions of the audience on gendered question-asking behavior. Fears of backlash are driven by concerns of potential negative judgments from others, leading to preemptively curbing public behavior (He et al., 2019; Moss-Racusin & Rudman, 2010; Rudman & Fairchild, 2004). In the realm of conferences, the extent to which women perceive a conference to have a sexist climate relates to how silenced they feel (Biggs et al., 2018). To test the role of fears of backlash in question-asking behavior, I manipulated fears of backlash by describing the audience as either a typical academic audience or a friendly audience. As in the previous studies, I expected that fears of backlash would be particularly salient for women. By reducing such concerns through a friendly audience, gender differences in question-asking intentions would be reduced. This study was preregistered (https://aspredicted.org/ys5fy.pdf).

Method

Participants. Participants were 792 workers (393 men, 399 women) from Prolific. On average, participants were 25 years old (*SD* = 7 years). In total, 6 participants were excluded for being outside the gender binary and 3 for not ever having enrolled in college.

Design. This experiment was a 2 (gender: men, women) by 2 (audience: typical, friendly) design.

Procedure. Participants read the same scenario as in Study 4 with added text for the audience manipulation. Participants in the friendly condition read, *"You look around at the audience, and it seems friendly and open."* Participants in the control condition read, *"You look around at the audience, and it seems like a typical academic audience."* Participants then answered the three question asking intention questions (a = .89), how many questions they could ask without being considered too demanding, how much time they could take without being seen as too pushy, and the fear of backlash scale (a = .92; Rudman & Fairchild, 2004). Participants also answered the extent to which they thought the audience was friendly, hostile, knowledgeable, typical, supportive, and intimidating on 7-point Likert scales (1 = Not at all; 7 = Extremely). Lastly, participants completed an attention check on the type of audience they read about and answered demographic questions.

Results

Manipulation Check. Participants in the friendly condition were more likely to miss the manipulation check than participants in the typical condition, b = 0.09, $t(790)$ = 3.83, $p < .001$, *95% CI* [0.04, 0.13]. Manipulation check failures did not differ by gender or the interaction with gender (ps > .83). Participants in the friendly condition would mistakenly select "typical" (N = 59) and several chose "hostile" (N = 4). I did not expect so many participants to be unable to correctly identify their condition at the end of the study (N = 92). Though I did not pre-register mis-identifying their condition as an exclusion criteria, I chose to exclude these participants to limit the sample to the participants who were cognizant of the key manipulation for the highest data quality possible.

Among the remaining sample, the manipulation acted in the ways I expected. Participants in the friendly condition rated the audience as friendlier, $t(678)$ = 6.65, $p < .001$, d = 0.50, *95%*

CI [0.35, 0.65], $M_{typical}$ = 4.53, $SD_{typical}$ = 1.08, $M_{friendly}$ = 5.09, $SD_{friendly}$ = 1.16, and more supportive, *t*(679) = 3.72, *p* < .001, *d* = 0.28, *95% CI* [0.13, 0.43], $M_{typical}$ = 4.45, $SD_{typical}$ = 1.13, $M_{friendly}$ = 4.78, $SD_{friendly}$ = 1.21, than participants in the typical condition. Additionally, participants in the typical condition rated the audience as more typical than participants in the friendly condition, *t*(684) = 4.31, *p* < .001, *d* = 0.33, *95% CI* [0.18, 0.48] , $M_{typical}$ = 5.15, $SD_{typical}$ = 1.22, $M_{friendly}$ = 4.75, $SD_{friendly}$ = 1.27.

Analytic Approach. Most analyses used 2 (gender: men, women) by 2 (audience: typical, friendly) ANOVAs. Simple effects tests were checked if interactions were significant. Reports for numbers of questions and amount of time were analyzed using negative binomial regressions to account for the count data. If negative binomial regressions failed to converge, ANOVAs were used.

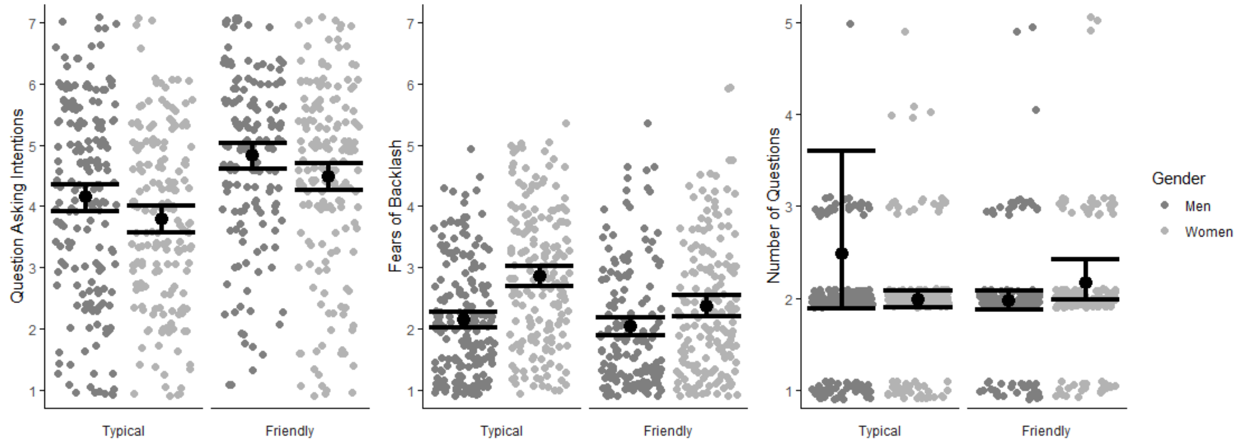
Question Asking Intentions. For question-asking intentions, there was a main effect of gender such that men (*M* = 4.49, *SD* = 1.53) had higher question-asking intentions than women (*M* = 4.17, *SD* = 1.49), *F*(1, 696) = 9.77, *p* = .002, *d* = 0.24, *95% CI* [0.09, 0.38], and a main effect of audience such that participants had higher question-asking intentions in the friendly audience (*M* = 4.65, *SD* = 1.46) compared to the control audience (*M* = 4.02, *SD* = 1.50), *F*(1, 696) = 37.48, *p* < .001, *d* = 0.46, *95% CI* [0.31, 0.61] (see Figure 3). However, there was no significant interaction between gender and audience-type, *F*(1, 696) = 0.02, *p* = .895.

Fears of Backlash. There was a significant gender by audience interaction, *F* (1, 696) = 6.10, *p* = .014, with a main effect of gender, *F*(1, 696) = 48.04, *p* < .001, *d* = 0.52, *95% CI* [0.37, 0.67], and a main effect of audience, *F*(1, 696) = 14.67, *p* < .001, *d* = 0.29, *95% CI* [0.14, 0.44]. Women had greater fears of backlash in typical audiences (*M* = 2.82, *SD* = 1.10) than friendly audiences (*M* = 2.42, *SD* = 1.11), *b* = 0.49, *t*(696) = 4.46, *p* < .001, whereas men did not differ in their fears of backlash by typical (*M* = 2.19, *SD* = 1.00) or friendly (*M* = 2.06, *SD* = 1.05) audiences, *b* = 0.10, *t*(696) = 0.93, *p* = .351. Women reported greater fears of backlash than men in both the typical audience, *b* = -0.34, *t*(696) = 4.46, *p* < .001, and the friendly audience, *b* = -0.73, *t*(696) = -6.73, *p* < .001.

Number of Questions. Gendered fears of backlash impacted how many questions participants felt they could ask before receiving backlash. There was a significant interaction, *b* = 0.31, *Z* = 2.65, *p* = .008, *95% CI* [0.08, 0.54], with no main effects of gender, *b* = -0.07, *Z* = -1.12, *p* = .264, *95% CI* [-0.18, 0.05], or audience, *b* = -0.07, *Z* = -1.19, *p* = .233, *95% CI* [-0.19, 0.05] (see Figure 3). Men thought they could ask more questions in the typical audience (*M* = 2.48, *SD* = 6.90) than the friendly audience (*M* = 1.97, *SD* = 0.67), *b* = 0.23, *Z* = 2.72, *p* = .007, whereas the number of questions women thought they could ask did not significantly differ by typical (*M* = 1.99, *SD* = 0.65) or friendly (*M* = 2.18, *SD* = 1.69) audience, *b* = -0.09, *Z* = -1.03, *p* = .305. Additionally, in the typical condition, men thought they could ask more questions before receiving backlash than women, *b* = 0.22, *Z* = 2.77, *p* = .006, with no significant gender differences in the friendly condition, *b* = -0.09, *Z* = -1.04, *p* = .298.

Amount of Time. However, gendered fears of backlash did not translate into estimates of how much time participants felt they could take before receiving backlash. There was no effect of gender, *b* = -0.03, *Z* = -0.54, *p* = .588, *95% CI* [-0.14, 0.08], audience, *b* = 0.10, *Z* = 1.72, *p* = .085, *95% CI* [-0.01, 0.21], or interaction, *b* = 0.18, *Z* = 1.55, *p* = .122, *95% CI* [-0.05, 0.40], on the amount of time participants anticipated they could take before experiencing backlash.

Figure 3. Responses for question-asking intentions, fears of backlash, and the number of questions that could be asked before experiencing backlash by audience and gender.



Moderated Mediation. A strength of the design of this study is that I directly manipulated the hypothesized mechanism (fears of backlash) through the climate of the audience. To test the causal relationship between fears of backlash on question-asking intentions and the gendered experiences of fears of backlash, I tested a moderated mediation with 10,000 bootstrapped samples with the LAVAAN package in R (Rosseel, 2012). The path model tested the effect of audience (-0.5 = typical, 0.5 = friendly) on question-asking intentions through fears of backlash. It specifically tested how the relationship between the audience and fears of backlash was moderated by gender (-0.5 = men, 0.5 = women).

Overall, fears of backlash mediated the effect of the audience on question-asking intentions (see Table 15). I found men and women's fears of backlash due to the audience type were significantly different from each other, *b* = -0.39, *SE* = 0.16, *Z* = -2.46, *p* = .014, *95% CI* [-0.70, -0.08]. Next, I looked at the effects by gender within the moderated mediation model. While typical audiences elicited fears of backlash for women, *b* = -0.49, *SE* = 0.12, *Z* = -4.26, *p* < .001, *95% CI* [-0.72, -0.26], they did not significantly elicit fears of backlash for men, *b* = -0.10, *SE* = 0.11, *Z* = -1.00, *p* = .319, *95% CI* [-0.31, 0.11]. Additionally, there was a significant indirect path of audience on question-asking intentions through fears of backlash for women, *b* = 0.16, *SE* = 0.05, *Z* = 3.30, *p* = .001, *95% CI* [0.08, 0.26], that was not present for men, *b* = 0.03, *SE* = 0.04, *Z* = 0.97, *p* = .334, *95% CI* [-0.03, 0.10].

Table 15. Path analysis estimates for the effect of question-asking intentions via fears of backlash.

Effect	Estimate	SE	Z	p	95% CI
Audience → Fear of Backlash (a)	-0.30	0.08	-3.80	<.001	-0.45, -0.14
Gender → Fear of Backlash (z)	0.53	0.08	6.84	<.001	0.38, 0.68
Fear of Backlash → Question Asking Intentions (b)	-0.32	0.05	-6.18	<.001	-0.42, -0.22
Total Effect (c)	-0.16	0.05	-3.23	.001	-0.26, -0.07
Direct Effect (c')	0.58	0.11	5.33	<.001	0.37, 0.79
Indirect Effect	0.10	0.03	3.02	.003	0.04, 0.17

Exploratory Audience Perceptions.

Knowledgeable. In exploratory analyses, we analyzed condition and gender differences on how the audiences were perceived. Friendly audiences (M = 5.12, SD = 1.18) were seen as less knowledgeable than typical audiences (M = 5.33, SD = 1.11), $F(1, 696) = 6.47$, $p = .011$, d = 0.19, *95% CI* [0.04, 0.34]. There was no significant effect of gender, $F(1, 696) = 3.68$, $p = .055$, d = 0.15, *95% CI* [0.00, 0.29], or its interaction with audience type, $F(1, 696) = 0.29$, $p = .592$.

Intimidating. Women (M = 4.21, SD = 1.74) perceived audiences to be more intimidating than men (M = 3.24, SD = 1.66), $F(1, 696) = 67.95$, $p < .001$, d = 0.62, *95% CI* [0.47, 0.77]. As would be expected, typical audiences (M = 4.15, SD = 1.71) were seen as more intimidating than friendly audiences (M = 3.31, SD = 1.73), $F(1, 696) = 60.59$, $p < .001$, d = 0.59, *95% CI* [0.44, 0.74]. There was no gender by audience interaction, $F(1, 696) = 1.64$, $p = .201$.

Hostile. There was a significant interaction, $F(1, 696) = 5.38$, $p = .021$, with a main effect of gender, $F(1, 696) = 9.35$, $p = .002$, d = 0.23, *95% CI* [0.08, 0.38], and a main effect of audience, $F(1, 696) = 31.08$, $p < .001$, d = 0.42, *95% CI* [0.27, 0.57]. Both men, b = 0.27, $t(696) = 2.27$, $p = .024$, and women, b = 0.65, $t(696) = 5.60$, $p < .001$, thought typical audiences were more hostile than friendly audiences. However, women (M = 2.51, SD = 1.24) thought typical audiences were more hostile than men (M = 2.12, SD = 1.12), b = -0.44, $t(696) = -3.82$, $p < .001$, while there were no significant differences between men (M = 1.88, SD = 1.10) and women (M = 1.96, SD = 1.06) in ratings of hostility within friendly audiences, b = -0.05, $t(696) = -0.42$, p = .673.

Interim Discussion

In Study 5, I found friendly audiences mitigated gender differences in fears of backlash. However, the friendly audience led men, in addition to women, to have higher question-asking intentions. Additionally, it led men to report that they could ask more questions before experiencing backlash, whereas the audience-type did not significantly impact how many questions women thought they could ask. While friendly audiences helped reduce women's negative experiences, it emboldened men, resulting in a persistent gender gap in participation. Additionally, I found that fears of backlash mediated the effect of the audience on question-asking intentions, which is causal support for the impact of fears of backlash on question-asking intentions. This experience also differed by gender, providing further support that fears of backlash are particularly salient to women in Q&A sessions.

Women rated all audiences as being more intimidating and typical audiences as more hostile than men. These differing perceptions of the same audience by gender could contribute some explanation as to why women are more reluctant to contribute to Q&A sessions and their fears of backlash. Fears of backlash are motivated by perceptions of external actors (Moss-Racusin & Rudman, 2010; Rudman & Fairchild, 2004). This study found that changing the environment could help women be more likely to engage, but did not uniquely help women. The next study builds on this one by testing how the content of the question interplays with the tenor of the audience to impact gender differences in question-asking intentions and fears of backlash.

Study 6: Impact of Question Content

In Study 6, I built on Study 5 by testing the impact of not only the audience, but the content of the question on question-asking intentions. I tested a boundary condition of when gender differences emerge in Q&A participation by manipulating the content of the potential question to be asked. Fears of backlash are elicited by creating the opportunity for stereotype-incongruent behavior (Eagly & Karau, 2002; Heilman et al., 2004; Heilman, 2012; Rudman & Glick, 1999). When participating in Q&A sessions, asking questions can be perceived as gender-incongruent for women if the act of question asking is framed as challenging the speaker and presenting as disagreeable; by contrast, if the act of asking questions is framed as presenting as agreeable and helpful, then it might be perceived as stereotype-congruent for women. Additionally, by specifying both the audience and the content, I tested not just stereotype violations but also broader group norm violations as a source for fears of backlash.

I hypothesized that there would be lower likelihoods of participating and greater fears of backlash when the question is framed as a critical question compared to a praise and an extension. I also hypothesized that there would be lower likelihoods of participating and greater fears of backlash when the audience is critical as opposed to supportive and when their question was incongruent to the audience compared to congruent. This study was preregistered (https://aspredicted.org/av5jd.pdf).

Method

Participants. Participants were 680 academics (183 men, 497 women) recruited from a society listserv. To protect the anonymity of the participants, I did not ask specifics about other demographic identities but instead only if participants held other marginalized identities in academia. In this sample, 222 participants identified as People of Color and 152 participants identified as part of the LGBTQ community. Academics of all ranks were represented in this sample as well as some participants from other industries (Undergraduate = 61, Post-Bac = 23, Graduate Student = 306, Postdoctoral Scholar = 57, Adjunct or non-Tenure Track Faculty = 7, Assistant Professor =89, Associate Professor = 62, Professor = 55, Industry = 20)[7].

Design. This experiment was a 2 (gender: men, women) by 2 (audience: hostile, friendly) by 2 (question: critical, praise and an extension) design.

Procedure. Participants were instructed to imagine that they were at a conference session they found interesting, and after watching a talk had a question to share with the speaker. They were told the question was either a critical question or a praise and an extension. The audience was then described as being critical and challenging to the speakers or friendly and supportive.

---

[7] The demographic question did not include options for post-bac roles (e.g., lab managers, post-bacs, research coordinators, or for non-tenure track faculty). Participants wrote these options in as "other." It is possible that other participants were in these ranks but selected adjacent ranks instead).

Measures. Participants responded to the same questions as in Study 4 (question-asking intentions: a = .86; fears of backlash: a = .90). Participants also answered questions about their general perceptions of their participation in Q&A sessions as in Study 3.

Results

Manipulation Check. Participants in the friendly audience were more likely to select the wrong condition than the participants in the critical audience, b = 0.08, $t(678) = 2.73$, $p = .007$, *95% CI* [0.02, 0.14]. Participants mistakenly identified the audience as "typical" (N = 62) or "hostile" (N = 12). There were no gender differences in selecting the wrong condition or interactions (ps > .30). Participants in the critical question condition were more likely to select the wrong condition than the participants in the praise condition, b = -0.10, $t(677) = -2.88$, p = .004, *95% CI* [-0.17, -0.03]. Participants mistakenly identified the question type as a clarification question (N = 77) or praise (N = 47). There was not strong evidence for gender differences in question type errors, b = -0.07, $t(675) = -1.74$, $p = .082$, *95% CI* [-0.15, 0.01], with men selecting the incorrect condition more frequently than women. Unfortunately, a surprising number of academics could not identify one (N = 166) or both (N = 83) of their conditions. Like Study 5, I did not preregister excluding participants who failed the manipulation check, but I chose to exclude participants who failed the manipulation check from the analyses within the experiment to ensure we had the highest quality data for our analyses.

Among the remaining participants, the audience manipulation worked as expected. The friendly audience was rated as friendlier, $t(428) = 6.68$, $p < .001$, $d = 0.65$, *95% CI* [0.45, 0.85], $M_{critical} = 3.96$, $SD_{critical} = 1.29$, $M_{praise} = 4.74$, $SD_{praise} = 1.12$, and more supportive, $t(425)$ = 5.24, $p < .001$, $d = 0.51$, *95% CI* [0.31, 0.70], $M_{critical} = 3.84$, $SD_{critical} = 1.29$, $M_{praise} = 4.47$, $SD_{praise} = 1.20$, than the critical audience. Similarly, the critical audience was rated as more hostile, $t(427) = 5.34$, $p < .001$, $d = 0.52$, *95% CI* [0.32, 0.71], $M_{critical} = 3.04$, $SD_{critical} = 1.48$, $M_{praise} = 2.35$, $SD_{praise} = 1.20$, and intimidating, $t(414) = 2.53$, $p = .012$, $d = 0.24$, *95% CI* [0.05, 0.43], $M_{critical} = 4.46$, $SD_{critical} = 1.60$, $M_{praise} = 4.07$, $SD_{praise} = 1.65$, than the friendly audience. There was no significant differences on the more neutral, "typical," qualifier, $t(419) = 0.19$, p = .851, $d = 0.02$, *95% CI* [-0.17, 0.21], $M_{critical} = 4.61$, $SD_{critical} = 1.21$, $M_{praise} = 4.63$, SD_{praise} = 1.14.

Analytic Approach. Most analyses used 2 (gender: men, women) by the hypothesized condition variable – either 2 (audience: critical vs friendly) or 2 (question-type: critical vs praise) – ANCOVAs controlling for the condition variable not in the interaction. Simple effects tests were checked if interactions were significant. Reports for numbers of questions and amount of time were analyzed using negative binomial regressions to account for the count data. If negative binomial regressions failed to converge, ANOVAs were used. General attitudes towards Q&A sessions were tested with t-tests comparing responses by gender.

Question Asking Intentions.

Question Type. There was no effect of gender, $F(1, 425) = 1.44$, $p = .231$, $d = 0.13$, *95% CI* [-0.09, 0.35], question type, $F(1, 425) = 0.58$, $p = .446$, $d = 0.07$, *95% CI* [-0.12, 0.26], or their interaction, $F(1, 425) = 0.30$, $p = .584$, on question-asking intentions.

Audience Type. There was a main effect of audience type, $F(1, 425) = 25.81$, $p < .001$, d = 0.49, *95% CI* [0.30, 0.68], such that participants had higher question-asking intentions with a friendly audience (M = 4.55, SD = 1.27) compared to participants with a critical audience (M = 3.78, SD = 1.53). There was no effect of gender, $F(1, 425) = 1.44$, $p = .231$, $d = 0.13$, *95% CI* [-0.09, 0.35], or its interaction with audience type, $F(1, 425) = 0.66$, $p = .417$, on question-asking intentions.

Fears of Backlash.

Question Type. There was a main effect of gender such that women (M = 2.69, SD = 1.07) reported greater fears of backlash than men (M = 2.39, SD = 0.92), $F(1, 425) = 6.74$, p = .010, d = 0.29, *95% CI* [0.07, 0.51]. There was no effect of question-type, $F(1, 425) = 0.84$, p = .359, d = 0.09, *95% CI* [-0.10, 0.28], or its interaction with gender, $F(1, 425) = 2.62$, p = .107.

Audience Type. There was a main effect of gender such that women (M = 2.69, SD = 1.07) reported greater fears of backlash than men (M = 2.39, SD = 0.92), $F(1, 425) = 6.71$, p = .010, d = 0.29, *95% CI* [0.07, 0.51]. There was no effect of audience-type, $F(1, 425) = 0.38$, p = .539, d = 0.06, *95% CI* [-0.13, 0.25], or its interaction with gender, $F(1, 425) = 0.61$, p = .436.

Number of Questions.

Question Type. There was no significant effect of gender, $F(1, 418) = 2.18$, p = .140, d = 0.17, *95% CI* [-0.05, 0.39], question-type, $F(1, 418) = 0.09$, p = .760, d = 0.03, *95% CI* [-0.16, 0.22], or their interaction, $F(1, 418) > 0.01$, p = .946.

Audience Type. There was no significant effect of gender, $F(1, 418) = 2.19$, p = .140, d = 0.17, *95% CI* [-0.05, 0.39], audience-type, $F(1, 425) = 1.91$, p = .167, d = 0.13, *95% CI* [-0.06, 0.32], or their interaction, $F(1, 418) = 1.01$, p = .316.

Amount of Time.

Question Type. There was no effect of gender, b = -0.07, Z = -0.88, p = .379, *95% CI* [-0.24, 0.09], question-type, b = 0.12, Z = 1.39, p = .166, *95% CI* [-0.05, 0.28], or their interaction, b = -0.15, Z = -0.88, p = .379, *95% CI* [-0.48, 0.18].

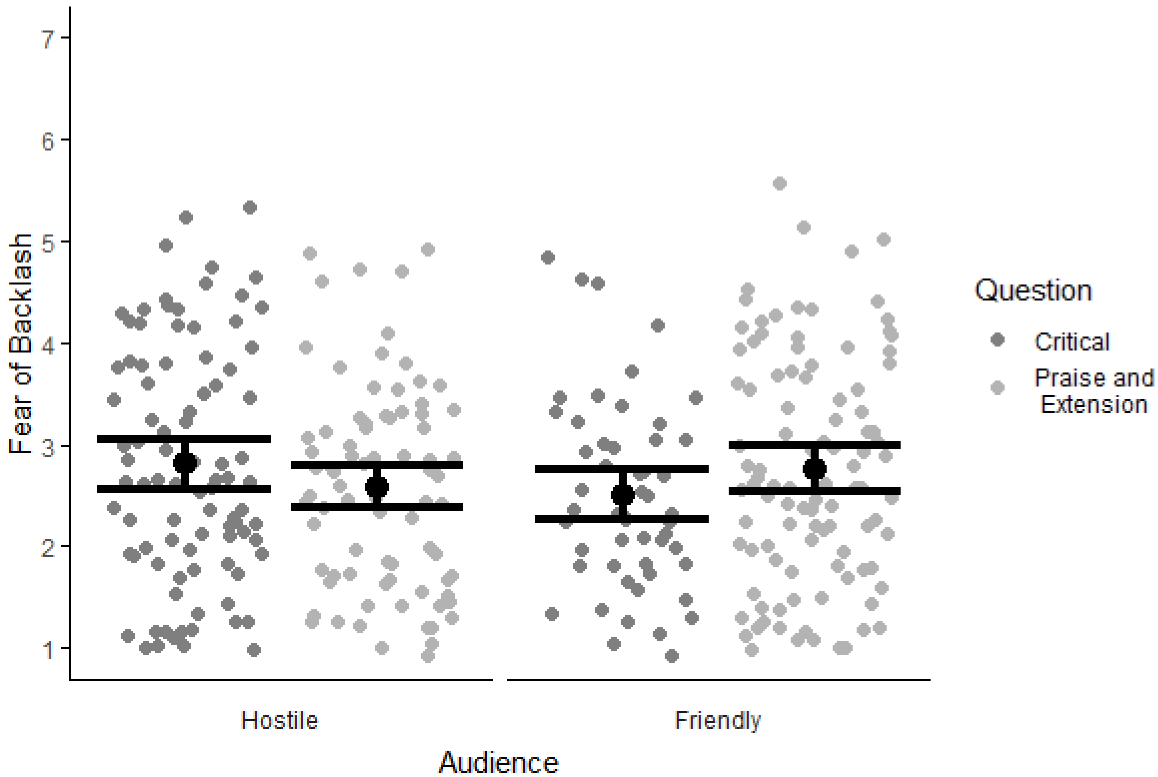
Audience Type. There was no effect of gender, b = -0.1=04, Z = -0.50, p = .617, *95% CI* [-0.21, 0.12], audience-type, b = -0.04, Z = -0.43, p = .670, *95% CI* [-0.20, 0.13], or their interaction, b = 0.32, Z = 1.90, p = .058, *95% CI* [-0.01, 0.65].

Exploratory Women-Only Analyses. Due to the oversampling of women in this study, I subsetted the data and analyzed interactions between audience conditions and question type conditions among only the women.

Question Asking Intentions. There was a significant main effect of audience type, $F(1, 321) = 24.58$, $p < .001$, d = 0.48, *95% CI* [0.29, 0.67], in which women had higher question-asking intentions when the audience was friendly (M = 4.53, SD = 1.26) as opposed to critical (M = 3.72, SD = 1.52). There was no effect of question type, $F(1, 321) = 0.81$, p = .370, d = 0.09, *95% CI* [-0.10, 0.28], or audience by question type interaction, $F(1, 321) = 0.68$, p = .412.

Fears of Backlash. There was a significant audience by question-type interaction, $F(1, 321) = 3.98$, p = .047, with no main effects of audience, (1, 321) = 0.08 p = .773, d = 0.03, *95% CI* [-0.19, 0.25], or question-type, $F(1, 321) < 0.01$, p = .960, d = 0.01, *95% CI* [-0.21, 0.22] (see Figure 4). Examining the simple effects, this difference-in-difference seems to be driven by women fearing more backlash for critical questions than praise and extensions, though the effect is weak, b = 0.32, $t(321) = 1.70$, p = .090 (ps for all other simple effects > .14).

Figure 4. Ratings of women's fears of backlash audience type by question type.



Number of Questions. There was no effect of audience, $F(1, 317) = 0.51$, $p = .478$, $d = 0.08$, *95% CI* [-0.14, 0.30], question type, $F(1, 317) = 0.02$, $p = .883$, $d = 0.02$, *95% CI* [-0.20, 0.24], or their interaction, $F(1, 317) = 0.04$, $p = .835$, on the perceived number of questions that can be asked before experiencing backlash.

Amount of Time. There was no effect of audience, $b = 0.14$, $Z = 1.65$, $p = .099$, *95% CI* [-0.03, 0.30], question type, $b = 0.14$, $Z = 0.32$, $p = .753$, *95% CI* [-0.14, 0.19], or their interaction, $b = -0.10$, $Z = -0.57$, $p = .572$, *95% CI* [-0.43, 0.23, on the perceived number of questions that can be asked before experiencing backlash.

Exploratory Analyses by Gender and Race.

Question-asking intentions. There was no effect of gender, $F(1, 426) = 0.74$, $p = .391$, $d = 0.10$, *95% CI* [-0.12, 0.32], being a Person of Color, $F(1, 426) = 0.08$, $p = .784$, $d = 0.03$, *95% CI* [-0.16, 0.22], or their interaction, $F(1, 426) = 2.02$, $p = .156$, on question-asking intentions.

Fears of backlash. Women reported higher fears of backlash than men, $F(1, 426) = 6.21$, $p = .013$, $d = 0.28$, *95% CI* [0.06, 0.50]. There was no effect of being a Person of Color, $F(1, 426) = 0.28$, $p = .597$, $d = 0.05$, *95% CI* [-0.14, 0.24], or its interaction with gender, $F(1, 426) < 0.01$, $p = .960$.

Number of questions. There were no significant main effects of gender, $F(1, 419) = 2.35$, $p = .126$, $d = 0.17$, *95% CI* [-0.05, 0.39]. There was no effect of being a Person of Color, $F(1, 419) = 3.76$, $p = .053$, $d = 0.19$, *95% CI* [0.00, 0.38], or its interaction with gender, $F(1, 419) = 3.29$, $p = .071$.

Amount of time. There was a significant main effect of being a PoC with PoC reporting they could take more time than White people, $b = 0.25$, $Z = 2.70$, $p = .007$, *95% CI* [0.07, 0.43].

There was no significant effect of gender, $b = -0.11$, $Z = -1.21$, $p = .227$, *95% CI* [-0.29, 0.07], or its interaction with being a PoC, $b = -0.33$, $Z = -1.84$, $p = .066$, *95% CI* [-0.70, 0.02].

Exploratory Analyses by Gender and Rank.

Question-asking intentions. There was no effect of gender, $b = -0.07$, $t(416) = -0.16$, $p = .874$, *95% CI* [-0.98, 0.83], rank, $b = 0.07$, $t(416) = 0.37$, $p = .714$, *95% CI* [-0.29, 0.43], or their interaction, $b = -0.01$, $t(416) = -0.13$, $p = .897$, *95% CI* [-0.22, 0.19].

Fears of backlash. There was no significant effects of gender, $b = 0.61$, $t(416) = 1.95$, $p = .052$, *95% CI* [-0.01, 1.22], rank, $b = 0.04$, $t(416) = 0.31$, $p = .759$, *95% CI* [-0.21, 0.28], or their interaction, $b = -0.09$, $t(416) = -1.23$, $p = .219$, *95% CI* [-0.22, 0.05].

Number of questions. Participants at lower ranks reported being able to ask more questions than participants at higher ranks, $b = -0.04$, $t(409) = -2.02$, $p = .044$, *95% CI* [-0.08, 0.00]. There was no effect of gender, $b = -0.06$, $t(409) = -0.34$, $p = .737$, *95% CI* [-0.38, 0.27], or its interaction with rank, $b = -0.01$, $t(409) = -0.28$, $p = .776$, *95% CI* [-0.08, 0.06].

Amount of time. There was no effect of gender, $b = -0.16$, $Z = -0.72$, $p = .473$, *95% CI* [-0.64, 0.30], rank, $b = -0.02$, $Z = -0.59$, $p = .556$, *95% CI* [-0.07, 0.04], or their interaction, $b = 0.02$, $Z = 0.47$, $p = .639$, *95% CI* [-0.08, 0.13].

General Attitudes towards Q&A Sessions. In general, women seem to enjoy the experience of attending conferences more than men. Women reported liking going to conference sessions more, $t(271) = 2.40$, $p = .017$, $d = 0.21$, *95% CI* [0.04, 0.38], $M_{men} = 3.93$, $SD_{men} = 0.90$, $M_{women} = 4.11$, $SD_{women} = 0.72$, and found talks to be more interesting, $t(291) = 3.05$, $p = .003$, $d = 0.26$, *95% CI* [0.09, 0.43], $M_{men} = 3.25$, $SD_{men} = 0.82$, $M_{women} = 3.46$, $SD_{women} = 0.72$. However, there were no significant differences in how good participants thought Q&A sessions are, $t(291) = 1.17$, $p = .244$, $d = 0.10$, *95% CI* [-0.07, 0.27], $M_{men} = 3.30$, $SD_{men} = 0.87$, $M_{women} = 3.38$, $SD_{women} = 0.76$.

When it comes to participating in Q&A sessions, women reported less comfort participating than men. Compared to men, women reported less comfort asking questions, $t(313) = 4.74$, $p < .001$, $d = 0.41$, *95% CI* [0.24, 0.58], $M_{men} = 3.93$, $SD_{men} = 1.74$, $M_{women} = 3.23$, $SD_{women} = 1.66$, sharing comments, $t(299) = 5.25$, $p < .001$, $d = 0.45$, *95% CI* [0.28, 0.63], $M_{men} = 3.62$, $SD_{men} = 1.74$, $M_{women} = 2.85$, $SD_{women} = 1.58$, or approaching speakers after the talks, $t(319) = 3.20$, $p = .002$, $d = 0.28$, *95% CI* [0.11, 0.45], $M_{men} = 4.55$, $SD_{men} = 1.83$, $M_{women} = 4.04$, $SD_{women} = 1.79$. Of note, while men reported asking questions more frequently than women, $t(308) = 2.86$, $p = .005$, $d = 0.25$, *95% CI* [0.08, 0.42], $M_{men} = 2.43$, $SD_{men} = 0.90$, $M_{women} = 2.21$, $SD_{women} = 0.84$, there were no significant gender differences in the frequency of holding back questions, $t(317) = 0.59$, $p = .559$, $d = 0.05$, *95% CI* [-0.12, 0.22], $M_{men} = 3.50$, $SD_{men} = 0.76$, $M_{women} = 3.54$, $SD_{women} = 0.73$.

Interim Discussion

In Study 6, I found more general effects than effects due to the manipulations, though participants responded as expected to the manipulation checks. Women were more likely to fear backlash for their participation compared to men, and this did not vary by question type or audience. When responding to general questions about Q&A participation, women report greater discomfort than men. Men reported that they ask questions more frequently, but there were no significant gender differences in the extent to which they hold back questions.

When only comparing women, women had greater question-asking intentions in friendly audiences compared to hostile audiences, replicating Study 5. Additionally, women's fears of backlash differed on the basis of question-type and the audience-type, suggesting that women are also attuned to how their words might come across.

Study 6 found self-reports of gendered participation by academics mirror data from conferences. While the type of question may play a minimal role in determining behavior, the type of audience may play a factor for academics in how they engage in Q&A spaces. As outside influences continue to impact Q&A participation, in the next study, I tested the impact of how public the structure of asking questions in the Q&A sessions impact participation. To improve the realism of the study, I tested this factor in a mock-conference situation and collected the questions that would be asked.

Study 7: Impact of Q&A Format

In Study 7, I manipulate fears of backlash to directly test the causal links between fears of backlash and Q&A participation. Past research has manipulated fears of backlash by altering the context of the interaction to be public or private (Buser et al., 2021). Because fears of backlash are definitionally socially related, fears of backlash should be eliminated when the entire interaction occurs anonymously. In this experiment, participants engaged in a mock-conference situation in which they were given the choice of one of four research presentations to watch. Participants were then randomly assigned for the Q&A context to be in-person, in identifiable chats, or anonymous chat. By including a condition in which participants asked questions identifiably but written, I can rule out the alternative explanation that the effects are driven by the difference between participating publicly versus anonymously rather than participating orally versus in written communication. I hypothesized that (1) women will ask fewer questions in general compared to men, and (2) gender differences will be highest when questions are posed publicly in front of a whole group as compared to anonymously to a speaker. This study was preregistered (https://aspredicted.org/9d2nx.pdf).

Method

Participants. Participants (*N* = 471, 160 men, 311 women) were recruited through an on-campus participant pool. Data collection for this study is ongoing. The analyses below are for the purpose of completing the dissertation and will be updated in future publications.

Design. This experiment was a 2 (gender: men, women) by 3 (audience: live, identifiable chat, anonymous chat) design.

Procedure. Participants first read a description of an undergraduate/graduate conference happening on campus and were asked to imagine that they chose to attend:

> *Every year, several departments at UC Berkeley organize day-long conferences for undergraduates to learn more about research their peers, and graduate students, and post docs worked on over the last year. During this event, students are placed in groups of four to present in a single session. Each student speaks for 5-8 minutes, leaving about 10 minutes for questions at the end. These sessions tend to be attended by 50-100 people on average ranging from undergraduates to professors.*
>
> *Imagine a conference like this occurred in your department and you decided to attend because some of the topics looked interesting to you and you are considering going to graduate school one day.*

Next, participants were given the option to watch one of four research talks, simulating the same kind of active choice participants at conferences make when choosing which session to attend. The talks were from a pool of presentations available online from conferences that occurred within the last year[8]. All talks were approximately the same length of time and were

[8] Two talks were longer than the desired length, so the speakers created abridged versions for the purpose of this study.

pretested to be similarly engaging and interesting. Participants were told that they would be asked questions about the talk after and would be encouraged to pay attention. After watching the video, participants were asked to summarize the talk to confirm that they watched it and were then given space for a private brainstorm session to write all of their reactions to the talks.

After, participants were given the structure of the Q&A, which was either live, through identifiable chats, or through anonymous chats with related pictures. The live condition read:

At this conference, questions are asked after all of the speakers have finished presenting. People interested in asking questions can raise their hands to be called on and ask their question to one of the speakers.

Participants in the identifiable chats condition read:

At this conference, questions are submitted via a phone app with questions attributed to your name and are visible to the entire audience. The speakers are then able to view the questions and answer them after all of the speakers have finished presenting.

Participants in the anonymous condition read:

At this conference, questions are submitted anonymously via a phone app and are visible to the entire audience. The speakers are then able to view the questions and answer them after all of the speakers have finished presenting.

Participants were able to view their brainstorm and were asked to write what they would say or write to the speaker to ask a question, or they were given the option to type "No questions" if they would not try or want to talk to the speaker. Participants responded to the same questions as in Study 1 (question-asking intentions: a = .88; fears of backlash: a = .92). Participants also answered questions about their reaction to the presentation.

Results

Manipulation Check. Participants in the identifiable chat conditions were more likely than participants in the anonymous chat condition, b = -0.10, Z = -2.62, p = .024, and the live condition, b = 0.14, Z = 3.54, p = .001, to miss the manipulation check. Participants in the chat condition mistakenly selected anonymous chat (N = 32) and a few said live (N = 4). Missing the manipulation check did not differ by gender or gender by condition interactions (ps > .18).

Analytic Approach. Most analyses used 2 (gender: men, women) by 3 (condition: live, chat, anonymous chat) ANOVAs. Simple effects tests were checked if interactions were significant. Reports for numbers of questions and amount of time were analyzed using negative binomial regressions to account for the count data. If negative binomial regressions failed to converge, ANOVAs were used.

Question Asking Intentions. For question-asking intentions, there was a main effect of gender such that men (M = 3.66, SD = 1.56) had higher question-asking intentions than women (M = 3.29, SD = 1.48), F(1, 464) = 6.45, p = .011. There was no significant effect of Q&A format, F(2, 464) = 2.20, p = .112, or gender by Q&A format interaction, F(2, 464) = 0.22, p = .801.

Fears of Backlash. Women (M = 2.48, SD = 1.05) reported greater fears of backlash compared to men (M = 2.48, SD = 1.09), F (1, 465) = 11.01, p < .001. There was no significant effect of Q&A format, F(2, 465) = 0.29, p = .750, or gender by Q&A format interaction, F(2, 463) = 1.29, p = .276.

Number of Questions. There was a significant main effect of gender, F(1, 455) = 5.17, p = .023, with women (M = 1.87, SD = 0.62) reporting they could ask more questions before experiencing backlash than men (M = 1.73, SD = 0.62). There was no effect of Q&A format,

$F(2, 455) = 0.03$, $p = .972$, or its interaction with gender, $F(2, 455) = 0.96$, $p = .385$ on the number of questions that could be asked.

Amount of Time. For the amount of time participants thought they could take, there was no significant effect of gender, $b = -0.10$, $Z = -0.72$, $p = .469$. There was an effect of Q&A format such that participants in the identifiable chat ($M = 52.98$, $SD = 44.89$) thought they could take more time than participants in the anonymous chat condition ($M = 44.66$, $SD = 38.11$), $b = -0.22$, $Z = -2.22$, $p = .026$. All other effects were nonsignificant (*p*s > .13).

Interim Discussion

As of the current data collection efforts, the overall gender effects emerged. That is, men reported greater question-asking intentions and women reported greater fears of backlash for their participation. However, there was no effect of how the Q&A period was structured on gendered participation or even participation more broadly. Of note, participants in the identifiable chat condition were more likely to misremember their condition than the other two conditions, in the direction of thinking their condition was anonymous chats. This could be evidence that the modality of the Q&A (i.e., chats) was more salient than how public their participation would be.

As of this writing, I am not interpreting the effects from this study definitively because I am missing a large amount of my projected sample size. If, after I complete data collection, I find no effects of Q&A format, there can be a number of interpretations of the null effects. Null effects could indicate that women do not modulate their participation based on the format of the Q&A session and that the observed behavioral differences could be due to other factors like who gets called on or who is able to get in line first. Fears of backlash would then be primarily driven by putting oneself out there in the first place as opposed to how public the forum is. It could be possible that women are more likely to have questions they would ask due to demand effects of being asked to write a question. It could also be possible that the brainstorming aspect of the design either gave women time to be more confident in their questions or men to consider their thoughts and self-censor more. Lastly, it is possible that the overall gender effect explains so much variance that it leaves less room for format-type to show effects.

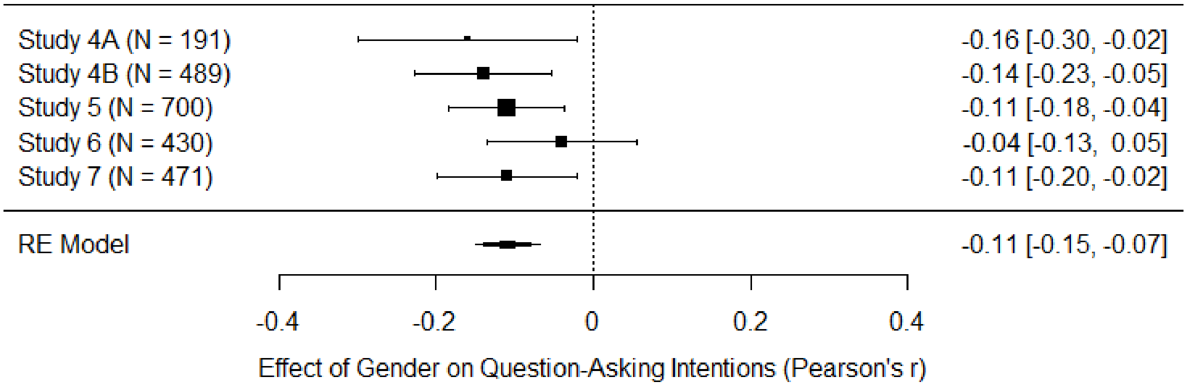
Internal Meta-Analysis

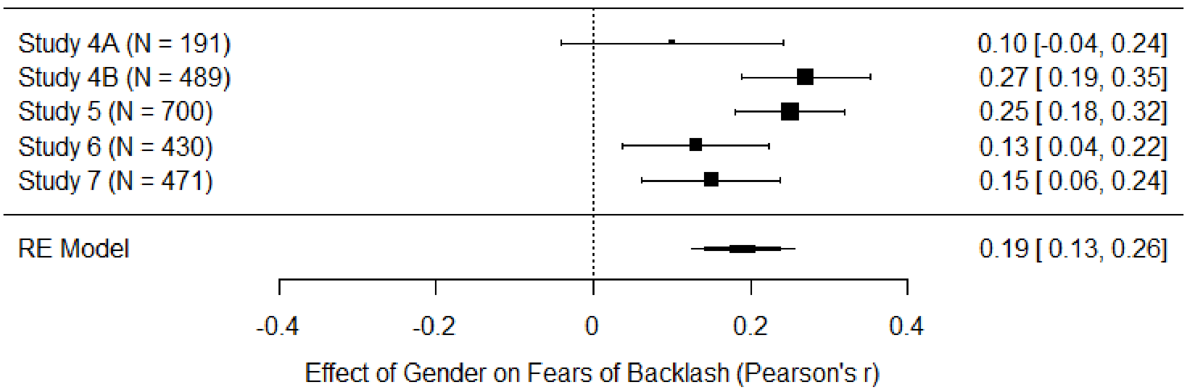
To aggregate the combined evidence across all of the reported studies, I conducted an internal meta-analysis on the effect size between gender and question-asking intention and fears of backlash. In total, the internal meta-analysis included the five effect size estimates from 2,623 participants (the five studies from this chapter: Studies 4 a/b - Study 7). There was a significant meta-analytic effect for question-asking intentions, $r = -0.11$, *95% CI* [-0.15, -0.07], $Z = -5.22$, $p < .001$, and for fears of backlash, $r = 0.19$, *95% CI* [0.13, 0.26], $Z = 5.74$, $p < .001$[9], indicating that across studies men indicated higher question-asking intentions and women indicated higher fears of backlash (see Figure 5). Of note, the estimates from Study 4A were similar to the observed effects from the other studies, just with a larger confidence interval around them. This is further evidence that Study 1A was simply too underpowered to observe the main effects. For limitations on internal meta-analyses, see Vosgerau et al. (2019).

---

[9] Because some backlash items were inadvertently dropped in Study 4A, analyses were also run only including the items present in Study 4A to ensure effects were not dependent on which items were included. Results were consistent with the full scale $r = 0.17$, *95% CI* [0.10, 0.24], $z = 4.72$, $p < .001$.

Figure 5. Forest plots of the effects of question-asking intentions and fears of backlash across the studies in Chapter 2.





Discussion

Across four studies, I found evidence for the role fears of backlash held by women have on question-asking intentions and the factors that impact their fears of backlash. In Study 4, when sample size was increased, men reported greater question-asking intentions and women reported greater fears of backlash. Further, fears of backlash explained the variance between gender and question-asking intentions. In Study 5, I tested the impact of a warmer environment in the audience on gendered participation. Women reported greater question-asking intentions and reduced fears of backlash with a friendlier audience. Further, the impact of fears of backlash caused by the audience on question-asking intentions was uniquely experienced by women. However, the environment did not reduce the gender gap in anticipated participation because it increased men's question-asking intentions and the number of questions they thought they could ask before experiencing backlash.

In Study 6, a sample of academic women reported greater fears of backlash compared to men, which did not vary significantly by the environment in the audience or the type of question that they would have asked. However, in analyses comparing effects only among women, women reported greater question-asking intentions in friendly audiences compared to critical audiences with no effect of question-type. In Study 7, I tested whether structural variation on how Q&A sessions were run would impact gendered participation. In general, men reported greater question-asking intentions and women reported greater fears of backlash for their participation. However, there was no effect of how the Q&A period was structured on gendered participation or even participation more broadly.

Taken together, these data suggest fears of backlash are a pervasive concern for women when they consider whether to participate in Q&A sessions. Rather than anxieties sparked by the kind of contribution women are considering making or how publicly the Q&A session is structured, women's fears of backlash seem to stem from the culture of the group in which they would be participating.

Of note, Study 6 replicated effects from Study 3, which found large gender effects on discomfort participating in Q&A sessions, yet no significant gender differences in the extent to which participants indicated that they hold back questions. That said, the same men also reported participating in Q&A sessions more than women and had greater question-asking intentions irrespective of the tenor of the audience or the type of question that would be asked. Future research is necessary to understand the discrepancy between gender differences in participation but not in holding back questions.

Limitations

A limitation of this work is that it relies on self-reported intentions of behaviors and simulations of behavior. However, our findings on gender differences in question-asking intentions are consistent with substantial behavioral evidence for observed gender differences in participation at conferences (Hinsley et al., 2017; Käfer et al., 2018; Pritchard et al., 2014; Schmidt & Davenport, 2017; Telis et al., 2019). By testing effects in this manner, I was able to more directly connect the psychological experience of participating in Q&A sessions with the commonly observed outcome of gender gaps in participation to try to understand why such gaps exist.

Additionally, I failed to find effects using some of our manipulations. Given our statistical methodology, an absence of observing an effect does not necessarily mean that there are statistically no differences between the groups. However, with sample sizes as large as the ones used in this chapter, it is likely that if there are true effects, they would be too small to be of practical interest.

Implications

Women have fears of backlash for participating in Q&A sessions, which are legitimate based on observations of how academic women are treated (Dupas et al., 2021; Gruber et al., 2021; Llorens et al., 2021; Shen, 2013). The present research found that women's greater fears do not stem from appearing more critical in Q&A sessions or from participating publicly in front of a large group. Rather, the main driver of backlash identified in these data was the climate of the audience. To the extent that women's fears of backlash are due to cultural aspects of the environment, simply telling women to be less anxious or more assertive will not reduce gender gaps in participation (Cheryan & Markus, 2020). Academics in academic spaces will need to consider what they can do to improve their environments and to create more welcoming and inclusive spaces for their women colleagues.

Chapter 3: Virtual Conferences

Introduction

Due to the onset of the COVID-19 pandemic, virtual conferences have become not only more popular and widespread, but necessary. Virtual conferences come with big hopes for inclusion in science. They have cheaper registration fees, require no travel or lodging costs, and often have events that allow for global participation. While these shifts made strides for international researchers and researchers with fewer resources, it is not yet known how the shift to a virtual format impacted the experience of conferences for women, and specifically, in Q&A sessions. In-person conferences show strong effects of men participating in Q&A sessions more than women across disciplines (Hinsley et al., 2017; Käfer et al., 2018; Pritchard et al., 2014; Schmidt & Davenport, 2017; Telis et al., 2019). It could be the case that the structural shifts in Q&A format successfully disrupt the masculine defaults present in live Q&A formats, creating a more welcoming environment for equitable participation by gender. However, it is also possible that the gendered power dynamics are so pervasive that disproportionate participation persists despite the changes in format. This chapter seeks to investigate the impact of virtual conferences on disproportionate participation in Q&A sessions by gender.

Impacts of Structural Shifts on Behavior

Structures and embedded systems have profound impacts on behaviors. People tend to act in ways that are consistent with and supported by the existing structures, called "channel factors" (Lewin, 1947; Ross & Nisbett, 1991). By altering structures within situations, it is possible to reduce friction and open channels to make different behavioral responses more likely (Lewin, 1952). For example, changing default options on retirement enrollment plans and appointments to receive flu shots so that the default is to be enrolled and have appointments increases participation in both programs (Carroll et al., 2009; Chapman et al., 2016; Leventhal et al., 1965). In this way, minor structural changes to the choice architecture can have meaningful impacts in how people engage in decision-making processes.

However, not all existing structures are as gender-neutral as the ones described above. One process by which fields and spaces stay majority-male is by relying on masculine defaults. Masculine defaults are aspects of systems that reward behaviors that are more associated with men and masculinity (Cheryan & Markus, 2020). For example, policies that require self-nominations for promotion or awards (Kang, 2014) or value individual contributions over collaboration and teamwork (Diekman et al., 2010), reward individualism and self-promotion, traits more commonly found among men (Cheryan & Markus, 2020; He et al., 2019; Moss-Racusin & Rudman, 2010). Women experience backlash for acting counter-stereotypically (Heilman, 2012; Rudman & Glick, 1999). Thus, masculine defaults put women in a bind: they must either conform to the masculine defaults and risk backlash, or act in more feminine ways and be less likely to have their contributions acknowledged or valued.

Virtual conferences provide a unique opportunity to consider alternative ways to organize Q&A sessions and to test the impacts of the structural shifts on participation. To the extent that women's participation is inhibited because the environment rewards more masculine forms of behavior (i.e., individualism, aggression, and competitiveness), gender-based disproportionate participation should be reduced. For example, in live Q&A sessions, individualism is rewarded when question-askers prioritize the answering of their own questions over what questions might be most informative for the full group. In contrast, in a virtual environment, time and space is less constrained due to the near infinite availability of chat space. Similarly, aggression is no longer rewarded by vying for space in line or being called on first (i.e., before time runs out) because whether a question gets asked is more within the agency of the question-asker and whether they choose to type it or not. In chat messages, question-asking becomes less visible of an action, which could mean that the utility and benefits of peacocking and participating to garner status could be reduced, simultaneously making it a more open and productive space for participation.

Gender and Virtual Communication

While the creation of online discussion presents itself as an egalitarian utopia, it could also be the case that the same gendered communication dynamics are reproduced in this new environment. Previous research on gender and online communication finds that men were more likely to participate in online forums and online political debates (Albrecht, 2006; Baek et al., 2012). Sociolinguistic theory posits that men communicate to build status while women communicate to build rapport; these dynamics were reproduced online (Gefen & Ridings, 2005). This effect was also not sensitive to gender base rates. In an undergraduate online discussion board for a psychology course where women outnumbered men 3:1, women were more likely to use hedges and express agreement whereas men were more likely to use authoritative language and express disagreement (Guiller & Durndell, 2006).

In a context similar to the present investigation, researchers investigated participation in Stack Overflow by gender. Stack Overflow is an online forum in which people can post questions related to computer coding and receive crowd sourced answers. In general, men were overrepresented in both question and answer posts and had higher status than women (Vasilescu et al., 2012). The authors argue that these gender differences could be due to the fact that status is driven through competitively earning prizes and the speed of response, and that men in that space do not consider sexism to be a problem. However, if women encountered more women when they first joined the platform, they engaged more quickly in the community (Ford et al., 2017).

In contrast, the volubility of communication by gender could be flipped in online support groups for health issues. In one study, women participated more in online cancer support groups compared to men (Ginossar, 2008). The researchers suggest that this effect is driven by the purposes of these groups, support conversations, which is more similar to the ways in which women communicate offline. However, a meta-analysis on support group communication found the support for this gendered effect to be mixed (Mo et al., 2009). In general, the existing research suggests that communication dynamics observed in in-person interactions are manifested virtually as well.

To date, work on virtual communication has primarily focused on online written communication. Little is known about the gendered experience of online video communication in terms of individual volubility and disproportionate participation. This chapter seeks to address this gap in the literature by testing for the impacts of gender communication in both written and video formats.

Competing Hypotheses

Between the structural shifts and pervasive gender-based status differentials in academia, two competing hypotheses are reasonable and plausible:

Hypothesis 1a: Disproportionate gender participation will be ameliorated in virtual environments.

Hypothesis 1b: Disproportionate gender participation will persist in virtual environments.

In this chapter, I investigate the extent to which there is evidence to support these hypotheses across several Q&A structures and contexts.

Present Studies

The aims of the present studies are to (1) test gender differences in Q&A behaviors across a variety of contexts to investigate the impacts of moving to a virtual environment on gender disparities in participation, and (2) test which structural factors can exacerbate or mitigate gender disparities in participation using variations in Q&A formats between and within conferences. Study 8 provides an initial test on the impacts of moving to a virtual environment on Q&A participation by gender and was used in an exploratory manner to determine which of the two competing hypotheses would be most likely. In Study 9, I tested the same conference as Study 1, only one year later. Though small, this conference provided the most direct test of the impacts of moving to a virtual format from a live format on gender disparities in participation by comparing gendered participation within the same organization.

In Study 10, the conference was majority women and participation was only available via text messaging. This allows for a test of whether gender disparities are mitigated in the circumstances theoretically most amenable to encourage participation from threatened groups. Numerical representation and opportunity space to engage are high while visibility is low. If gender disparities persist, this would be the most conservative test of the impacts of structural aspects. The conference tested in Study 11 was a large conference, functioning as the main conference for an entire field. The Q&A format varied widely across sessions within the conference, allowing for exploration of which circumstances are most likely to reduce (or exacerbate) gender disparities in participation.

Study 8: Exploratory Conference

Methods

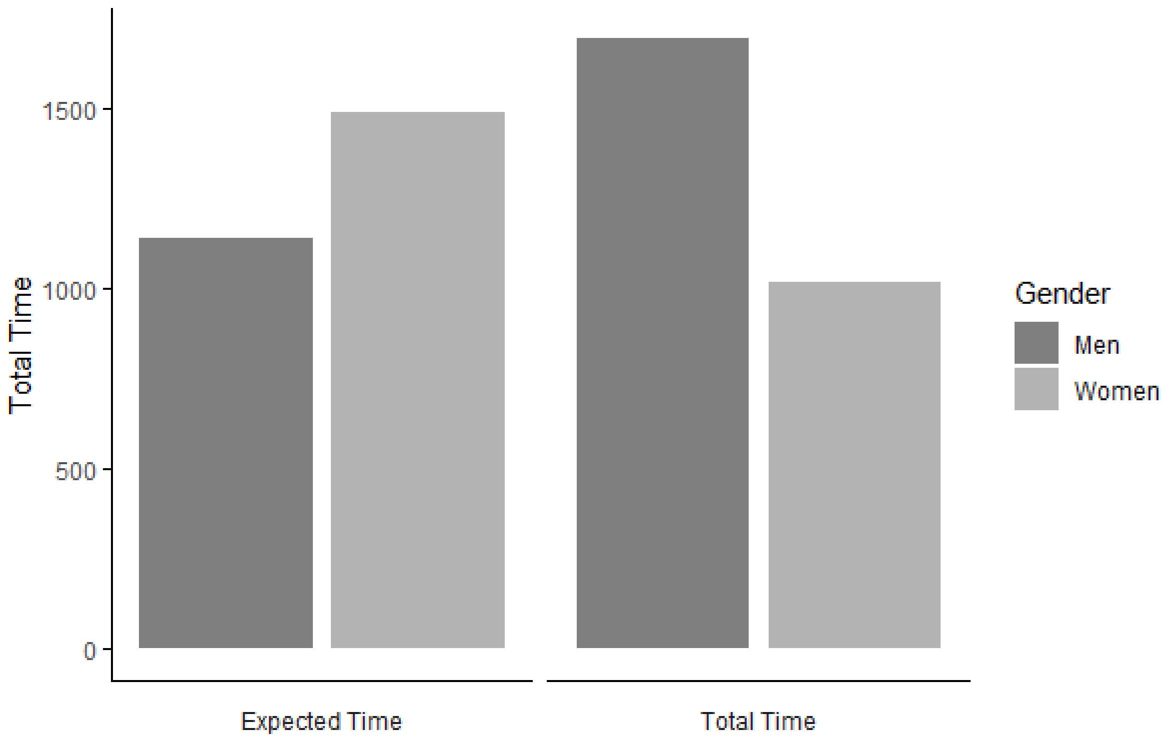
Data Source. This conference was a small, topical conference that varied from being single or dual track across three conference days. It was attended by 252 attendees (66 men and 103 women) and had 23 speakers (16 men and 7 women). The conference occurred on Zoom and was recorded and posted online. There were 53 Q&A interactions across 20 research talks. Several (4) research talks either did not have Q&A or had technology failures and were not recorded. Most questions were asked on video during the sessions, though some were asked via chat and answered on video. I was unable to collect the chat messages from the conference organizers, so Q&A interactions were only counted if they were answered on video. Further, six interactions were not included in analyses because the questions were asked via chat and were not attributed to the question-asker when answered over video.

Qualitative Coding. Two research assistants watched the Q&A sessions from each talk and coded Q&A behaviors according to the codebook from Study 1. Research assistants timed the length of the participant remarks and speaker responses. The recorded time from the two research assistants were considered discrepant if the difference was greater than 5 seconds (*weighted Kappa* = .91). If the difference was less than 5 seconds, the two measurements were averaged and rounded to the nearest second. If the difference was five seconds or more, a senior member of the team also timed the discrepant measurement and the original measurement closest to the third measurement was selected.

Results

There were no significant differences in propensity to initiate Q&A interactions (60% men, 40% women, χ^2 (1, N = 47) = 1.85, p = .174, d = 0.40, *95% CI* [-0.19, 0.99], or length of speaking time for questions, b = -0.12, Z = -0.58, p = .56 (M_{men} = 60.39, SD_{men} = 39.56; M_{women} = 53.42, SD_{women} = 36.65). Men took more total Q&A time than women compared to gender base rates of the conference (62% men, 38% women, χ^2 (1, N = 2706) = 196.26, p < .001, d = 0.56, *95% CI* [0.48, 0.64] (see Figure 6).

Figure 6. Proportion of participation by men and women in Q&A sessions. The first set of columns is the expected rate of participation given the base rates of attendees. The second set of columns is the total amount of time used by gender.



Interim Discussion

Though two of the analyses did not show statistically significant differences, likely due to the small sample size, all patterns of results were in the same direction. Not only did men participate more than women compared to what would be expected based on the gender base rates of attendees, men participated more than women outright despite being the numerical minority of attendees at the conference. Due to the consistent pattern of results, based on these exploratory analyses, I decided to preregister null hypothesis tests predicting men would disproportionately participate compared to gender base rates of attendees (https://osf.io/h7wfd).

Study 9: Virtual vs. In-Person Conference Direct Comparison

Methods

Data Source. This conference was a small, interdisciplinary conference that varied from being single or dual track across two conference days. It was attended by 172 attendees (104 men, 67 women, 1 person could not be identified). Of the conference speakers, 32% were female and 65% were male. The conference occurred on Zoom and was recorded and posted online. There were 34 Q&A interactions across 13 research talks. Two sessions were not coded because the posted recordings were cut before the Q&A occurred (N = 5 questions). Most questions were asked on video during the sessions, though some were asked via chat and answered on video. A unique benefit of this conference is that the previous year's live conference was also recorded and posted online, allowing for a clean test for how the move from live conferences to virtual conferences impacted Q&A participation by gender within roughly the same population (see Study 1).

Qualitative Coding. The live data was coded using the same codebook as Study 1 (ICC = .98).

Gender. Gender base rates for the conference were approximated by monitoring the participant list on the Zoom window in each session and recording their names. Gender was assessed by pronouns as part of their name on Zoom (5%) or on their personal websites (23%), gender presentation on Zoom video (19%), their Zoom profile pictures (36%), or their personal websites (13%), and their names (4%). The gender of the speakers and question askers from live Q&A sessions was coded by two raters based on their gender presentation (*Cohen's Kappa* = .85).

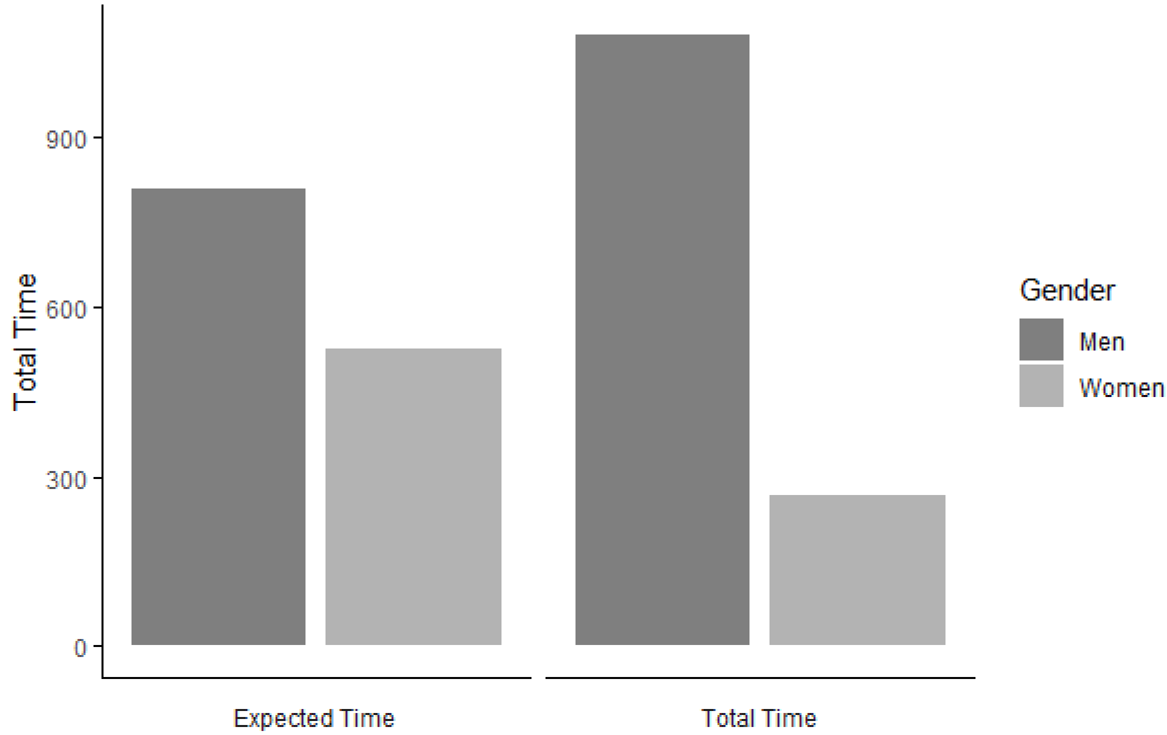
Results

Analytic Approach. First, gender differences in participation within the conference were tested using the same approach as Study 1. Then, to test the impact of switching from in-person to virtual formats, the gender differences from each format were compared.

Participation in the Virtual Conference. There were no significant gender differences in the propensity to initiate Q&A interactions(74% men, 26% women, χ^2 (1, N = 34) = 0.76, p = .383, d = 0.30, *95% CI* [-0.39, 0.99][10] or differences in lengths of speaking time, b = -0.41, Z = -1.14, p = .256, *95% CI* [-1.07, 0.34], (M_{men} = 44.08, SD_{men} = 45.10; M_{women} = 29.74, SD_{women} = 21.95. However, men occupied more of the total Q&A time compared to women (80% men, 20% women, (80% men, 20% women, χ^2 (1, N = 1370) = 128.37, p < .001, d = 0.64, *95% CI* [0.53, 0.75], see Figure 7).

[10] I attended this conference and recorded when each question was asked, by whom, and their gender presentation in addition to the coded data from the recorded videos. Of the five uncoded questions, 4 were asked by men and 1 was asked by a woman. This analysis was comparable when including the five uncoded questions, (71% men, 29% women, χ^2 (1, N = 38) = 0.93, p = .334, d = 0.31, *95% CI* [-0.34, 0.97].

Figure 7. Proportion of participation by men and women in Q&A sessions. The first set of columns is the expected rate of participation given the base rates of attendees. The second set of columns is the total amount of time used by gender.



Differences Between the In-Person and Virtual Conference. To test differences in disproportionate participation by gender between the in-person and virtual conferences, I compared the observed gender differences in participation for the in-person and virtual conferences (e.g., 47% in the virtual conference) to what would be expected based on the gender base rates of attendees (e.g, 21% in the virtual conference) using chi-square tests. There were no significant differences in men's disproportionate participation between the in-person and virtual conferences in propensity to ask questions, (χ^2 (1, *N* = 50) < 0.01, *p* > .999, *d* = 0.00, *95% CI* [-0.56, 0.56])[11] and total floor time (χ^2 (1, *N* = 3621) = 1.12, *p* = .290, *d* = 0.04, *95% CI* [-0.03, 0.10]).

Differences in gender differences in speaking time by live and virtual conferences were analyzed with a negative binomial regression including gender (effect coded: -0.5 = men, 0.5 = women), conference type (effect coded: -0.5 = in-person, 0.5 = virtual), and their interaction. There was an effect of gender such that across both types of conferences men spoke for a longer amount of time compared to women, *b* = -0.55, *Z* = -2.61, *p* = .009, *95% CI* [-0.94, -0.12]. There was no effect of conference type, *b* = -0.36, *Z* = -1.73, *p* = .084, *95% CI* [-0.76, 0.05], or a moderation of the gender difference by conference type, *b* = 0.28, *Z* = 0.67, *p* = .501, *95% CI* [-0.55, 1.10].

Interim Discussion

[11] This analysis was also comparable when including the five uncoded questions, (χ^2 (1, *N* = 52) < 0.01, *p* < .999, *d* = 0.00, *95% CI* [-0.55, 0.55]).

The key contribution of this study is the ability to compare gendered participation within the same conference in in-person and virtual formats one year apart. In both years, individually, men took more of the Q&A session time than women compared to what would be expected by the gender base rates of attendees. Analyses comparing the two conferences found that gender differences in participation or speaking time did not significantly vary, suggesting that virtual conferences with video Q&A sessions might not provide any different (or protective) environment for women compared to in-person conferences.

Study 10: Chat-Only Participation

Methods

Data Source. This conference was a conference for a subdiscipline and was attended by 3,888 attendees. All presentations were pre-recorded and available on the conference webpage. Some pre-recordings were scheduled during the conference period for collective watching (73 sessions), and others were available after the conference for asynchronous viewing. Because less emphasis is placed on engagement with poster sessions and asynchronous sessions compared to the “live” symposiums, only the “live” symposiums were analyzed. On the page for each presentation, there was a side-panel with a discussion thread specifically for Q&As and a discussion thread for general chats. Chat messages from the conference were web-scraped from the conference web portal one week after the official conference period ended. Both the Q&A window and chat window were checked for Q&A interactions[12]. Across the conference, there were 750 Q&A interactions[13] (446 at live research sessions) and 591 general chats (22 Q&A interactions were sent via general chat but analyzed with the Q&A interactions).

Gender. Gender base rates for the conference were determined by self-reported gender collected at registration and provided by the conference organizers. The conference was attended by 31% men and 65% women. The names of all chat senders and presenters were recorded. The gender identities of the chat senders and presenters was determined by looking for their gender identity on professional or social media webpages. Self-identified pronouns were prioritized (78% of chat senders and presenters), followed by gender presentation in photos (21% of chat senders and presenters), and lastly by their name (1% of chat senders and presenters). We were unable to identify gender identities for 3 participants.

Status. Participant status was determined by identifying which career stage they were in at the time of the conference. Web sources with career stages and year markers were prioritized in searches (e.g, CVs, LinkedIn, news articles), but information on university directories, personal web pages, and Twitter were also checked absent timed information. Status was classified into 7 categories: undergraduate, post bac (e.g., lab managers, project coordinators), graduate student, post doc, assistant professor (also included adjuncts and lecturers[14]), associate professor, full professor, and industry. Status could not be identified for 48 chat senders and presenters (5% of the sample).

---

[12] An unknown number of sessions engaged in some chats within video calls of their sessions instead of the window in the conference platform. However, most sessions used the Q&A window to some degree, and on 7 sessions did not have any activity in the Q&A window.

[13] To keep our coding of behaviors consistent with the video coding, if a participant who was not a speaker on the project replied to a Q&A thread, the reply (as well as corresponding speaker responses) was added to a new row and treated as a new Q&A interaction (N = 120). This method was designed to best capture attendee participation.

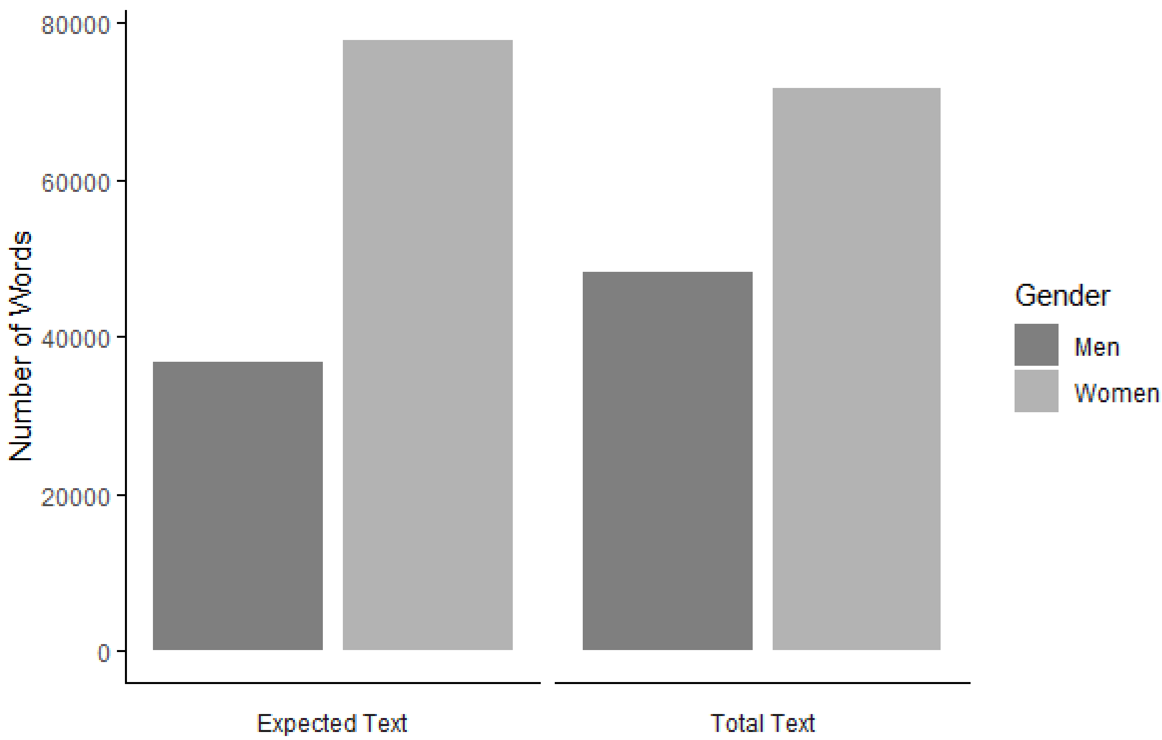
[14] Lecturer is a common title seen as equivalent to an assistant professor in countries outside of the US such as the UK and Europe.

Participation. To adapt to the completely text-based nature of Q&A interactions in this conference, participation was operationalized as engaging in Q&A chats with speakers by gender and the number of words used per initial question chat message and across all participant chats in the Q&A interaction.

Results

Participation. Men were more likely to initiate Q&A interactions compared to what would be expected by the gender base rates of conference attendees, χ^2 (1, N = 446) = 4.04, p = .044, d = 0.19, *95% CI* [0.00, 0.38]. There were no significant gender differences in total chat length, b = -0.06, Z = -1.04, p = .297, *95% CI* [-0.17, 0.05], (M_{men} = 279 words, SD_{men} = 179 words; M_{women} = 263 words, SD_{women} = 165 words). Across the conference, men provided a disproportionate amount of the total Q&A question text compared to gender base rates, 40% men, 59% women, χ^2 (1, N = 119918) = 1684.4, p < .001, d = 0.24, *95% CI* [0.23, 0.25] (see Figure 8).

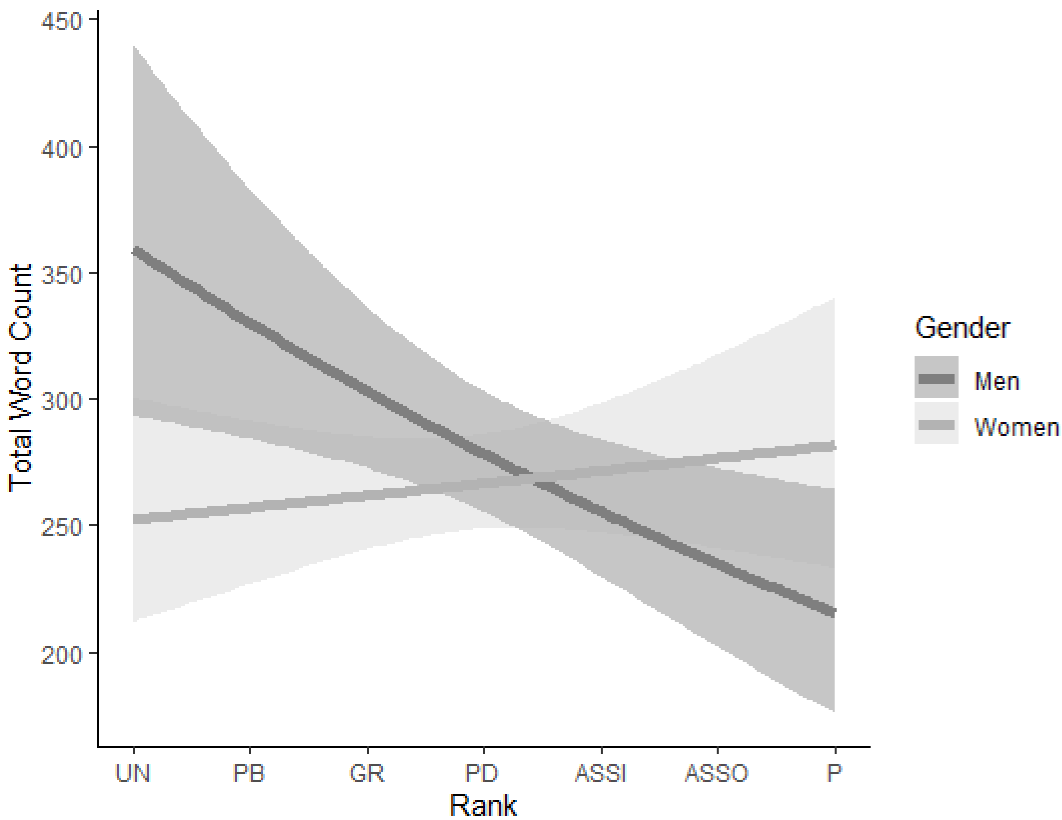
Figure 8. Proportion of participation by men and women in Q&A sessions. The first set of columns is the expected rate of participation given the base rates of attendees. The second set of columns is the total text used by gender.



Moderating Effects of Status. To test the moderating effects of status on gender differences in question length, status was treated in a continuous, linear manner. Question-askers who were identified as being in industry (N = 22) were excluded for analyses because it was unclear how they would fit into the academic hierarchy[15]. Status (centered) was entered into the negative binomial models as a predictor and interaction term with gender (effect-coded).

There was a significant gender by status interaction for the total text used in the Q&A interaction, $b = 0.11$, $Z = 2.52$, $p = .012$, *95% CI* [0.02, 0.19], with no main effect of gender, $b = -0.05$, $Z = -0.81$, $p = .419$, *95% CI* [-0.16, 0.06], or rank, $b = -0.03$, $Z = -1.65$, $p = .099$, *95% CI* [-0.08, 0.01]. Upon evaluating the simple effects, the shape of the interaction indicates gender differences among lower status positions in which men write more than women, and gender differences are ameliorated at higher status levels (see Figure 9). It is worth noting that the raw values for almost all of the status levels except graduate students show directionally that women had longer chat messages than men, suggesting that this interaction could be driven by the particularly large gender differences at that level (see Table 16). In a model testing gender differences in total text length for only graduate students, men wrote more than women, $b = -0.24$, $Z = -3.04$, $p = .002$, *95% CI* [-0.39, -0.09].

Figure 9. Fitted slopes for total word count of each Q&A interaction by gender and status.



---

[15] Arguably, adjunct professors and other non-tenure track faculty roles hold a lower status than tenure-track or tenured faculty in the university setting. Models were tested in which non-tenure-track faculty were ranked between post docs and assistant professors, and results were consistent to when non-tenure-track faculty were grouped with assistant professors. Given how few non-tenure-track faculty there were, we report them as in the same rank as assistant professors in the main text. See OSF for models with non-tenure track faculty as separate (https://osf.io/ebdgy/).

Table 16. Simple effects test within model testing gender by status interaction on total length of participant text in Q&A interactions for Study 10. Table displays *N*s, raw means and standard deviations, and simple effects within the context of the full model for each categorized career stage. Estimates are on the log (and not the response) scale.

	Men			Women						*95%CI*	
	N	*M*	*SD*	*N*	*M*	*SD*	*b*	*Z*	*p*	lower	upper
Undergrad	4	251	171	6	278	132	-0.36	-2.61	.009	-0.62	-0.09
Post Bac	4	166	93	1	317	--	-0.25	-2.51	.012	-0.45	-0.06
Grad Student	78	330	223	138	260	145	-0.14	-2.08	.038	-0.28	-0.01
Post Doc	28	266	106	28	246	153	-0.04	-0.66	.512	-0.15	0.07
Assistant Professor	33	232	122	59	294	224	0.07	0.97	.334	-0.07	0.21
Associate Professor	12	188	72	20	225	120	0.18	1.70	.090	-0.03	0.38
Professor	14	247	140	8	328	204	0.28	2.00	.045	0.00	0.56

Interim Discussion

Study 10 represented what could be arguably the best chance for women to reach equitable participation in Q&A sessions. Conference attendance and conference speakers had women in the numerical majority. Additionally, to the extent that the reason for women not participating is due to the stress of using collective space to visibly ask a question, these fears would be ameliorated in a completely virtual setting. Yet, men asked disproportionately more questions, and wrote disproportionately more of the question words across the conference.

Academic status moderated total engagement in the Q&A interaction. In particular, at the graduate student level, men wrote more than women across Q&A interactions, and this effect diminished as status increased. Here, it was not that women wrote particularly short messages; they were not holding back. Rather, male grad students wrote much more than any other group. While some may be concerned that gender gaps in Q&A participation are simply reflective of men being overrepresented in higher status academic positions, this finding adds to the evidence that does not find this to be the case (Hinsley et al., 2017). If anything, the raw means suggest that women directionally engaged more in the Q&A interactions compared to men, which would suggest that there could be some benefits of a virtual format to women that were largely overshadowed by the disproportionate participation of male grad students.

To further investigate the conditions under which women may be more likely to participate, we used, in Study 11, data from a large conference with video and text components. I tested differences between how Q&A sessions were formatted that varied in how public the participation would be (e.g., chat participation compared to video participation) and the availability for question asking (e.g., opportunities to ask questions between sessions compared to all questions at the end). Because the formats varied within one conference, I could test the impact of various formats on gendered participation within the same population.

Study 11: Participation Across Q&A Formats

Methods

Data Source. The conference was a large conference and served as the main conference for a field. It was attended by over 10,000 attendees. My intention was to use a web scraper to collect the names of all of the attendees and use an algorithm to approximate their likely gender based on their first names. There was an issue with the web scraper that was not recognized until after the conference closed in which only the first twenty attendees of each session were recorded (N = 3696). It is unclear how representative this set of attendees is of the entire conference. For this study, I was unable to test how representation differs from base rates. Instead, I test what factors predict participation.

Some symposiums were presented live with a live audience, and some were pre-recorded and available for asynchronous viewing. During live sessions, attendees were encouraged to only use the chat box on the conference webpage as opposed to within the Zoom window in order to allow future watchers to be able to see the conversations. Both the video recordings and chat messages were checked for Q&A interactions and were coded separately.

The conference included several types of events including research symposia, professional development workshops, award ceremonies, discussions, and networking events. In order to be comparable to the other conferences, only the research symposiums in which speakers presented novel research with time allotted for a Q&A session (either between talks or at the end of all of the talks) were coded and included in the analyses. Across the 1576 conference events, 861 events met our criteria for being research talks, and of those, 109 were live and included Q&A sessions.

The Q&A format varied by symposia as determined by each symposium organizer. Across the included symposia, there were a total of 601 Q&A interactions. Some symposia had Q&A between each talk (64%), while others saved all Q&A for the end (36%). Symposia differed between whether Q&A interactions were only initiated live (21%), only via chat (22%), or included both formats (55%).

All chat messages sent through the conference portal were web-scraped during the last week the conference was available at the end of the three-month conference period. A total of 14,311 chats were sent across all conference events. Only chats sent during live sessions with Q&A sessions (543 chats) were analyzed.

Q&A Session Coding. Live Q&A interactions were coded using the same codebook used in Study 1 (ICC = .99). Other data about each session were recorded, including aspects of how the Q&A sessions were structured (e.g., if there was time for questions after each talk or if all questions were at the end; if questions were only sent via chat, asked via video, or both were encouraged; whether the question was sent via chat or video; and from which division the talk was).

Gender. The gender of the speakers and question-askers from live Q&A sessions was coded by two raters based on their gender presentation (*Cohen's Kappa* = .84).

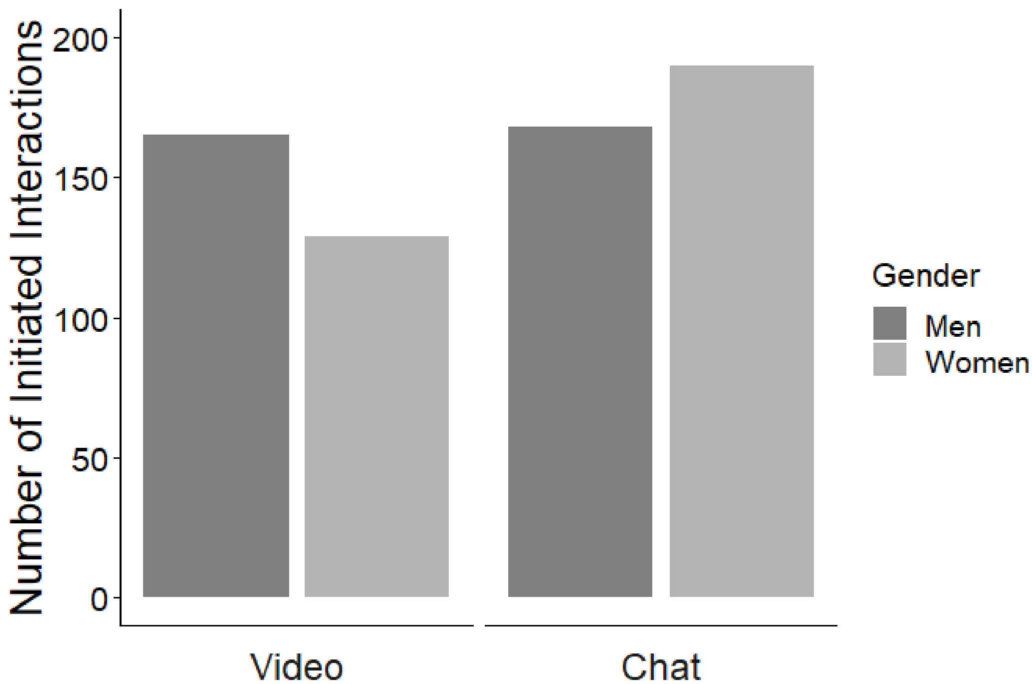
Results

Analytic Approach. Effects of structural aspects of Q&A sessions on gendered participation were tested on how they predict the gender of the question-asker and the quantity of participation (i.e., length of time or number of chat words). Gender of the question-asker was predicted using multi-level logistic regressions, and the quantity of participation was predicted using multi-level negative binomial regressions, each nested by symposium. Tested moderators include: whether Q&A was asked intermittently in the session or at the end effect coded (-0.5 =

end, 0.5 = middle), whether the question was asked via video or chat effect coded (-0.5 = video, 0.5 = chat), whether the format of the Q&A was in video, chat, or a combination of the two reference coded (in video as the reference group[16]), and the subfield (N = 3) associated with the session effect coded (-0.5 = not in that subfield, 0.5 = in that subfield)[17]. For the number of chat words, the moderator for the format of the Q&A only included in chat or a combination of video and chat effect coded (-0.5 = chat, 0.5 = combination of video and chat) because chat questions were not a structurally endorsed mode for question asking (excludes 18 questions). In the participation analyses, gender effect coded (-0.5 = men, 0.5 = women) was included as an interaction term for all of the moderators to test for gender differences in their impacts. A model with each division was also run to explore how micro-cultures impacted gendered participation. These models did not converge with a nested data structure, so they were run without a multilevel framework.

Participation. In general, women asked a higher proportion of the chat messages (53%) than video messages (44%), χ^2 (1, N = 652) = 5.10, p = .024, d = 0.25, *95% CI* [0.03, 0.47] (see Figure 10).

Figure 10. Number of initiated Q&A interactions by asking mode and gender.



However, counter to expectations, no structural aspects of Q&A sessions predicted the gender of the question-asker (see Table 17). Analyzing participation by division, four divisions specifically saw gendered participation: three in which women were more likely to participate

[16] In video was chosen as the reference group because it compares the chat features to the status quo.

[17] At this conference, sessions were associated with various divisions (N = 21) that group presentations by topical similarity. Sessions could be associated with up to 7 divisions, and on average were associated with 1.4 divisions. Ostensibly, sessions within a division would have similar group culture and gender representation. Divisions were too fine-grained for the structural analyses, so we moved up to the level of subfield. Each division is associated with one of three subfields, which vary by broad orientation towards research and serve as a proxy for likely gender representation in the session. Due to sessions being associated with multiple divisions, it was possible for a session to be in multiple subfields.

and one in which women were less likely to participate. These effects held while controlling for the structural predictors (see Table 18)[18].

Table 17. Likelihood of initiating a Q&A interaction by gender (men coded as 0, women coded as 1) predicted by each structural factor.

					95% CI	
	b	*odds ratio*	*Z*	*p*	*lower*	*upper*
Intercept	-0.14	0.87	-0.39	.698	-0.87	0.58
Middle vs End	0.45	1.57	1.41	.160	-0.19	1.10
Chat vs Video	0.28	1.32	1.31	.191	-0.14	0.70
Q&A Format: Chat	-0.65	0.52	-1.01	.313	-1.96	0.61
Q&A Format: Both	-0.10	0.90	-0.25	.805	-0.89	0.69
Subfield 1	-0.17	0.84	-0.38	.704	-1.05	0.73
Subfield 2	-0.24	0.79	-0.51	.610	-1.17	0.70
Subfield 3	-0.44	0.64	-1.21	.226	-1.19	0.29

Variance explained by random intercept for symposium session: 24%

[18] Some structural effects emerge in this model that were not present in the nested model. These significant effects are not being interpreted because the nested structure explained 24% of the variance in the model. It is likely that these effects are explained by within-session variance as opposed to between-session variance.

Table 18. Likelihood of initiating a Q&A interaction by gender (men coded as 0, women coded as 1) predicted by each significant division controlling for each structural factor.

					95% CI	
	b	*odds ratio*	*Z*	*p*	*lower*	*upper*
Intercept	-0.23	0.79	-1.06	.288	-0.66	0.19
Division 2	0.89	2.44	2.56	.011	0.22	1.59
Division 15	-0.79	0.45	-3.12	.002	-1.29	-0.30
Division 18	2.27	9.68	2.81	.005	0.86	4.19
Division 19	2.02	7.54	4.98	<.001	1.28	2.89
Middle vs End	0.46	1.58	2.39	.017	0.08	0.84
Chat vs Video	0.41	1.51	2.18	.030	0.04	0.78
Q&A Format: Chat	-0.25	0.78	-0.67	.503	-1.00	0.49
Q&A Format: Both	0.00	1.00	0.00	.999	-0.49	0.49

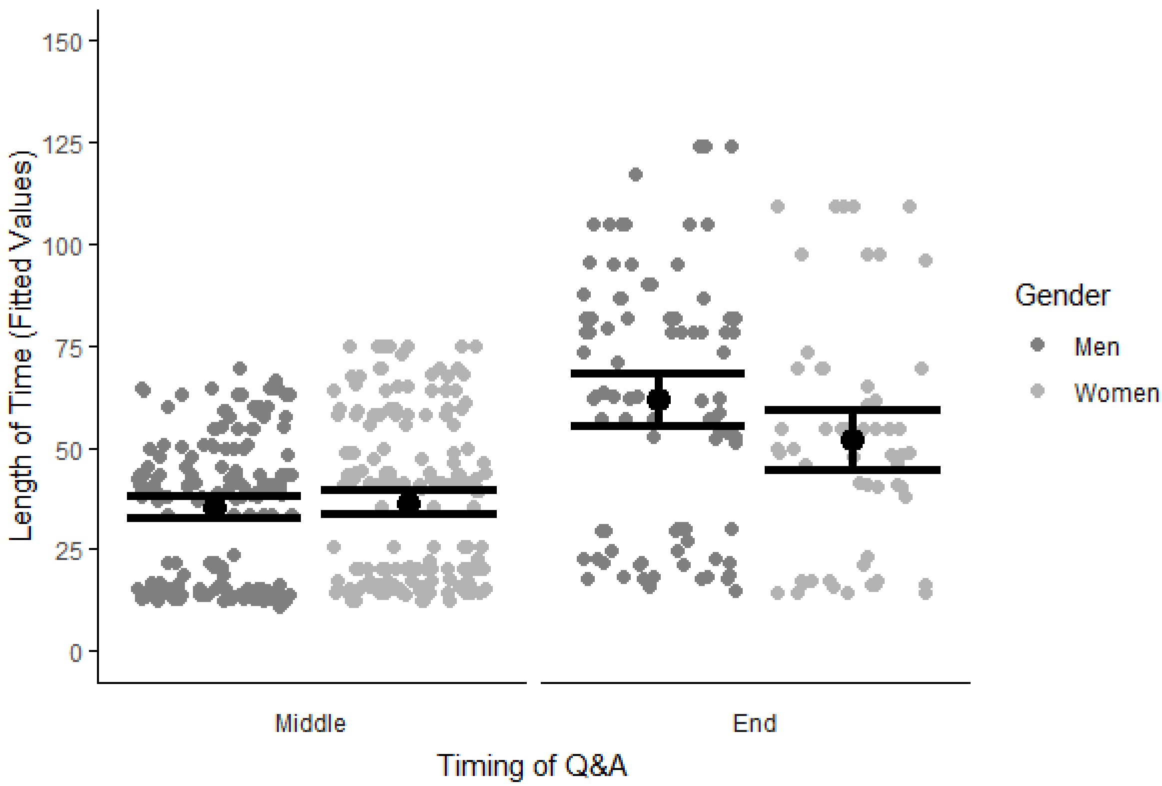
Length of Speaking Time. There were no significant gender differences between men (M = 46 seconds, SD = 46 seconds) and women (M = 41 seconds, SD = 40 seconds) on how long participants talked during the Q&A sessions in the video portions of the Q&A without controlling for the structural moderators, b = -0.06, Z = -0.82, p = .41, *95% CI* [-0.22, -0.09]. Of the structural moderators, one was significantly moderated by gender: whether questions were asked intermittently throughout the session versus all questions at the end, b = 0.38, Z = 2.35, p = .019 (see Table 19). There were no significant main effects of gender, b = -0.01, Z = -0.04, p = .970, or for if the questions were throughout the session versus all questions at the end, b = -0.19, Z = -1.87, p = .061. Reviewing the pattern of effects, men asked longer questions when all questions were saved for the end of session compared to when questions were throughout the session, b = -0.38, Z = -3.02, p = .003, whereas women did not change the length of their questions based on when they occurred, b = 0.00, Z = -0.03, p = .979 (see Figure 11).

Table 19. Length of speaking time predicted by gender, each structural factor, and gender's interaction with each structural factor.
Note: Unable to estimate 95% CIs for this model.

	b	*odds ratio*	*Z*	*p*
Intercept	3.69	40.04	32.00	<.001
Gender	-0.01	0.99	-0.04	.970
Middle vs End	-0.19	0.83	-1.87	.061
Chat vs Video	-1.06	0.35	-13.51	<.001
Q&A Format: Chat	-0.41	0.66	-1.81	.070
Q&A Format: Both	-0.26	0.77	-2.10	.036
Subfield 1	0.06	1.06	0.44	.660
Subfield 2	-0.04	0.96	-0.30	.766
Subfield 3	0.07	1.07	0.60	.549
Gender X Middle vs End	0.38	1.46	2.35	.019
Gender X Chat vs Video	0.01	1.01	0.10	.922
Gender X Q&A Format: Chat	0.09	1.09	0.28	.789
Gender X Q&A Format: Both	-0.11	0.90	-0.57	.571
Gender X Subfield 1	-0.11	0.90	-0.55	.583
Gender X Subfield 2	-0.08	0.92	-0.35	.729
Gender X Subfield 3	0.07	1.07	0.44	.663

Variance explained by random intercept for symposium session: 8%

Figure 11. Total word count of each Q&A interaction by gender and whether the question was asked in a session where questions were asked intermittently or at the end using the fitted values.



When analyzing gendered differences in question length by division, six divisions showed interactions that did not fall cleanly along subfield lines (see OSF: https://osf.io/bszef/). Similar to what was seen with overall participation, micro-cultures within particular topics may have more of an influence over how question-askers participate.

Length of Initial Question Text. Absent moderators, there was a significant effect of gender on word length of question text, b = -0.18, Z = -2.68, p = .007, *95% CI* [-0.31, -0.05], with men (M = 239 words, SD = 155 words) writing longer chat questions than women (M = 195 words, SD = 106 words). Including all of the moderators and their interactions with gender into the model, there were no interactions between gender and the moderators, indicating that there were not any structural or cultural effects that predicted when gender differences in chat length emerge (see Table 20). Additionally, including all of the structural factors washed out the gender effects, suggesting that the structural predictors could be explaining some of the same variance as gender. In the analysis testing gender effects within each division, only one division had a significant gender by division effect. Given that it was only one division of 21, it could be possible that this effect was a false positive and the micro culture of divisions had little effect on chat question length.

Table 20. Length of question text predicted by gender, each structural factor, and gender's interaction with each structural factor.

					95% CI	
	b	*odds ratio*	*Z*	*p*	*lower*	*upper*
Intercept	5.43	228.15	55.83	<.001	5.25	5.63
Gender	0.10	1.11	0.54	.589	-0.26	0.48
Middle vs End	-0.16	0.85	-1.83	.064	-0.33	0.01
Chat vs Video	-0.12	0.89	-0.71	.479	-0.47	0.20
Q&A Format	0.06	1.06	0.58	.561	-0.16	0.26
Subfield 1	0.01	1.01	0.11	.915	-0.19	0.22
Subfield 2	-0.27	0.76	-2.51	.012	-0.48	-0.05
Subfield 3	-0.07	0.93	-0.75	.455	-0.24	0.11
Gender X Middle vs End	0.00	1.00	-0.02	.986	-0.32	0.31
Gender X Chat vs Video	-0.39	0.68	-1.15	.252	-1.06	0.25
Gender X Q&A Format	-0.30	0.74	-1.57	.116	-0.67	0.07
Gender X Subfield 1	0.13	1.14	0.71	.476	-0.23	0.50
Gender X Subfield 2	0.37	1.45	1.86	.063	-0.02	0.77
Gender X Subfield 3	0.02	1.02	0.09	.926	-0.31	0.34

Variance explained by random intercept for symposium session: 1%

Interim Discussion

Counter to my predictions, there were minimal to no effects of changes in Q&A formats on gendered participation. There were no gender differences on speaking time of questions, replicating effects seen in live conferences (Study 2). However, men wrote longer questions in chats compared to women, replicating Study 10. Of note, within-session effects explained almost a quarter of the variance in whether there would be gendered participation, but did not explain variance in models predicting speaking length or number of words. It appears that some divisions had either particularly inhibitory or encouraging effects on whether women participated, though not the length of their participation. This may be some evidence that immediate environments may play an integral role in women's decision-making process for whether to participate. Without gender base rates, it's impossible to distinguish if this effect is a result of women being more likely to attend particular sessions. Should this effect be driven by selection effects, it would be worth considering why women are more comfortable attending particular sessions and how that relates to the culture within that particular research topic.

Discussion

Across three studies, we found men used more of the collective Q&A time and space at conferences compared to women, and that in general, gendered participation did not vary by Q&A format in a fourth study. In Study 8, men used more of the collective time than women as compared to what would be expected by base rates and more than women outright even though they were in the numeric minority. In Study 9, men's disproportionate participation did not significantly differ between in-person and virtual formats within the same conference on either side of the start of the pandemic. In Study 10, within an all-chat participation format, men asked a disproportionate number of questions. Additionally, greater verbosity was moderated by status, such that graduate students had large gender differences that were not observed at higher ranks. In Study 11, alterations to how the Q&A sessions were structured had minimal impacts on gender differences in Q&A participation. However, there was some evidence that environmental effects due to division-level differences could have particularly inhibitory or encouraging effects on women's participation.

The findings that men take disproportionate Q&A space is consistent with past research studying Q&A sessions at live conferences (Hinsley et al., 2017; Käfer et al., 2018; Pritchard et al., 2014; Schmidt & Davenport, 2017; Telis et al., 2019). Gender effects across modalities appear to be pervasive. Past work has also found that men have greater volubility at higher status levels whereas women do not change their participation by status (Brescoll, 2011). In contrast, this study found that men at lower statuses had more volubility that was not seen at higher levels. It could be the case that academic men have learned to step back or, more cynically, were less likely to engage in the virtual modality in general.

Limitations

A key limitation of this work is that it is observational. Certain conference attendees could choose to attend conferences with particular formats which may impact gender differences in participation outcomes. The relative consistency between the virtual and in-person effects suggest that this may not be too much of a concern. Additionally, the conclusions drawn from these data are limited by the conferences that were studied. I did my best to sample conferences with varying formats, sizes, lengths, and gender representations of attendees to allow these results to be as generalizable as possible. Coding conferences is time-intensive and access can be difficult. More work should be done to examine virtual conferences across fields and formats to test the robustness of these effects.

Implications

With the onslaught of virtual conferences at the start of the COVID-19 pandemic came excitement about the benefit of increased access for people restricted by distance, finances, home obligations, health and disability, etc. One might hope or expect that shifts in access to conferences would also translate to changes not just in attendance, but also participation. In the present data, the experience of virtual conferences was not markedly different from the experience of in-person conferences. Men continued to take disproportionate amounts of Q&A time and space compared to what would be expected by base rates. In the largest observed conference, some smaller divisions were deterministic in who participated by gender. It could be the case that the experience of gender gaps in participation is not impacted by structures, but due to deep-seeded cultural norms and expectations. More work is necessary to understand the gendered experiences and perceptions of Q&A sessions and how they translate to these behavioral outcomes.

General Discussion

Across the 11 studies in this dissertation, I investigated gendered participation in Q&A sessions with a multi-method approach using a combination of observational field data, experiments, and self-reports. In Chapter 1, I found gender differences in Q&A participation in live conferences were primarily driven by men being more likely to initiate Q&A interactions and not in other behaviors. Additionally, women reported feeling less comfortable than men participating in Q&A sessions. In Chapter 2, I found men reported greater question-asking intentions and women reported greater fears of backlash using hypothetical vignettes and experiments. In Chapter 3, I found men's disproportionate participation in virtual conferences persisted across all Q&A formats. Across studies, gendered experiences of Q&A sessions were pervasive, underscoring that inclusivity in Q&A sessions is a concrete issue for academia to address.

Men's disproportionate participation in Q&A sessions relative to the gender base rates of attendees was consistent across the observational field studies. My results are consistent with past research studying conference Q&A sessions (Hinsley et al., 2017; Käfer et al., 2018; Pritchard et al., 2014; Telis et al., 2019), and extends this work by testing differences in how gendered participation emerges with greater granularity and extends the breadth of contexts tested. Even in the face of factors that might be predicted to mitigate gendered participation, gender gaps in participation were persistent. Men's disproportionate participation was persistent regardless of whether women were in the numeric majority or minority. Participation gaps also persisted in in-person and virtual environments as well as when questions were asked in chats and videos. Though men self-report that they hold back questions to make space for others, as a collective, they reliably continue to take disproportionate space.

Across six samples, women consistently reported having greater fears of backlash for their participation in Q&A sessions compared to men. Additionally, fears of backlash explained some of the relationship between gender and question-asking intentions. These results are consistent with existing research on fears of backlash finding that fears of backlash result in reducing volubility and taking less space (Amanatullah & Morris, 2010; Brescoll, 2011; Moss-Racusin & Rudman, 2010).

An avenue by which fears of backlash could theoretically be reduced is through Q&A sessions where questions are submitted through chat platforms. Participation through chat is less publicly visible, which theoretically should reduce the salience of judgment from others. However, in the conference we studied that had chat-based Q&A sessions, men's disproportionate participation remained present. It is worth noting that the effect size of the disproportionate participation was the smallest in this conference. Additionally, in Study 11, where participants could attend sessions that varied in format, women contributed a greater proportion of the chat questions than the video questions. Data collection remains ongoing for the experiment that directly tests the effect of having chat-based Q&A sessions. Those data will help explain the role of chat messages in how they impact fears of backlash, if at all, when data collection is complete.

The data from this dissertation points to audience perceptions as one of the stronger contributing factors to gender gaps in participation. Women view typical and hostile audiences more negatively than men do. Negative perceptions of the audience influenced their anticipated behavior, similar to responses to conferences (Biggs et al., 2018). To the extent that the driving factor is due to a group culture, small, subtle alterations may not be sufficient to increase women’s participation in Q&A sessions. Broader cultural shifts require an intentional coordinated effort across interpersonal and structural levels (Markus & Kitayama, 2010).

Power differentials by gender in academia are pervasive and demonstrated with Q&A sessions. Demonstrating knowledge is akin to promoting one’s status in academia, and Q&A sessions provide a forum for public self-promotion in this way (Walker et al., 2010). To the extent women have fears of backlash either through witnessing instances of backlash or through other idiosyncratic sources, women stay silent, maintaining the gender hierarchy status quo (Phelan & Rudman, 2010). Ensuring women feel comfortable participating in public spaces should be a priority to increase the inclusivity of academia. This dissertation identifies Q&A sessions as one such space.

References

Aguillon, S. M., Siegmund, G.-F., Petipas, R. H., Drake, A. G., Cotner, S., & Ballen, C. J. (2020). Gender differences in student participation in an active-learning classroom. *CBE—Life Sciences Education*, *19*(2), ar12. https://doi.org/10.1187/cbe.19-03-0048

Akinola, M., Martin, A. E., & Phillips, K. W. (2018). To Delegate or Not to Delegate: Gender Differences in affective associations and behavioral responses to delegation. *Academy of Management Journal*, *61*(4), 1467–1491. https://doi.org/10.5465/amj.2016.0662

Albrecht, S. (2006). Whose voice is heard in online deliberation?: A study of participation and representation in political debates on the internet. *Information, Communication & Society*, *9*(1), 62–82. https://doi.org/10.1080/13691180500519548

Amanatullah, E. T., & Morris, M. W. (2010). Negotiating gender roles: Gender differences in assertive negotiating are mediated by women's fear of backlash and attenuated when negotiating on behalf of others. *Journal of Personality and Social Psychology*, *98*(2), 256–267. https://doi.org/10.1037/a0017094

Amanatullah, E. T., & Tinsley, C. H. (2013). Punishing female negotiators for asserting too much…or not enough: Exploring why advocacy moderates backlash against assertive female negotiators. *Organizational Behavior and Human Decision Processes*, *120*(1), 110–122. https://doi.org/10.1016/j.obhdp.2012.03.006

Anderson, K. J., & Leaper, C. (1998). Meta-analyses of gender effects on conversational interruption: Who, what, when, where, and how. *Sex Roles*, *39*(3), 225–252. https://doi.org/10.1023/A:1018802521676

Babcock, L., Recalde, M. P., Vesterlund, L., & Weingart, L. (2017). Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review*, *107*(3), 714–747. https://doi.org/10.1257/aer.20141734

Baek, Y. M., Wojcieszak, M., & Delli Carpini, M. X. (2012). Online versus face-to-face deliberation: Who? Why? What? With what effects? *New Media & Society*, *14*(3), 363–383. https://doi.org/10.1177/1461444811413191

Baker, T. L., & Vélez, W. (1996). Access to and opportunity in postsecondary education in the United States: A review. *Sociology of Education*, *69*, 82–101. https://doi.org/10.2307/3108457

Biggs, J., Hawley, P. H., & Biernat, M. (2018). The academic conference as a chilly climate for women: Effects of gender representation on experiences of sexism, coping responses, and career intentions. *Sex Roles*, *78*(5), 394–408. https://doi.org/10.1007/s11199-017-0800-9

Bowles, H. R., Babcock, L., & Lai, L. (2007). Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational Behavior and Human Decision Processes*, *103*(1), 84–103. https://doi.org/10.1016/j.obhdp.2006.09.001

Bowles, H. R., & Flynn, F. (2010). Gender and persistence in negotiation: A dyadic perspective. *Academy of Management Journal*, *53*(4), 769–787. https://doi.org/10.5465/amj.2010.52814595

Brescoll, V. L. (2011). Who takes the floor and why: Gender, power, and volubility in organizations. *Administrative Science Quarterly*, *56*(4), 622–641. https://doi.org/10.1177/0001839212439994

Brescoll, V. L., Okimoto, T. G., & Vial, A. C. (2018). You've come a long way…maybe: How moral emotions trigger backlash against women leaders. *Journal of Social Issues*, *74*(1), 144–164. https://doi.org/10.1111/josi.12261

Burgoon, J. K., & Dunbar, N. E. (2006). Nonverbal expressions of dominance and power in human relationships. *The Sage Handbook of Nonverbal Communication*, *2*, 279–297.

Burgoon, M., Dillard, J. P., & Ooran, N. E. (1983). Friendly or unfriendly persuasion. *Human Communication Research*, *10*(2), 283–294. https://doi.org/10.1111/j.1468-2958.1983.tb00018.x

Buser, T., Ranehill, E., & van Veldhuizen, R. (2021). Gender differences in willingness to compete: The role of public observability. *Journal of Economic Psychology*, *83*, 102366. https://doi.org/10.1016/j.joep.2021.102366

Carroll, G. D., Choi, J. J., Laibson, D., Madrian, B. C., & Metrick, A. (2009). Optimal defaults and active decisions. *The Quarterly Journal of Economics*, *124*(4), 1639–1674. https://doi.org/10.1162/qjec.2009.124.4.1639

Carter, A. J., Croft, A., Lukas, D., & Sandstrom, G. M. (2018). Women's visibility in academic seminars: Women ask fewer questions than men. *PLOS ONE*, *13*(9), e0202743. https://doi.org/10.1371/journal.pone.0202743

Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences*, *108*(8), 3157–3162. https://doi.org/10.1073/pnas.1014871108

Chapman, G. B., Li, M., Leventhal, H., & Leventhal, E. A. (2016). Default clinic appointments promote influenza vaccination uptake without a displacement effect. *Behavioral Science & Policy*, *2*(2), 40–50. https://doi.org/10.1353/bsp.2016.0014

Chen, J. M., & Moons, W. G. (2015). They won't listen to me: Anticipated power and women's disinterest in male-dominated domains. *Group Processes & Intergroup Relations*, *18*(1), 116–128. https://doi.org/10.1177/1368430214550340

Cheryan, S., & Markus, H. R. (2020). Masculine defaults: Identifying and mitigating hidden cultural biases. *Psychological Review*, *127*(6), 1022–1052. https://doi.org/10.1037/rev0000209

Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology*, *97*(6), 1045–1060.

Cialdini, R. B. (2001). The Science of Persuasion. *Scientific American*, *284*(2), 76–81.

Cook, J. E., Purdie-Vaughns, V., Garcia, J., & Cohen, G. L. (2012). Chronic threat and contingent belonging: Protective benefits of values affirmation on identity development. *Journal of Personality and Social Psychology*, *102*(3), 479–496. https://doi.org/10.1037/a0026312

Dardenne, B., Dumont, M., & Bollier, T. (2007). Insidious dangers of benevolent sexism: Consequences for women's performance. *Journal of Personality and Social Psychology*, *93*(5), 764–779. https://doi.org/10.1037/0022-3514.93.5.764

Diekman, A. B., Brown, E. R., Johnston, A. M., & Clark, E. K. (2010). Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychological Science*, *21*(8), 1051–1057. https://doi.org/10.1177/0956797610377342

Dupas, P., Modestino, A. S., Niederle, M., Wolfers, J., & Collective, T. S. D. (2021). *Gender and the Dynamics of Economics Seminars* (No. w28494). National Bureau of Economic Research. https://doi.org/10.3386/w28494

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*(3), 573–598. https://doi.org/10.1037/0033-295X.109.3.573

Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology*, *46*(4), 735–754. https://doi.org/10.1037/0022-3514.46.4.735

Fisher, A. N., Stinson, D. A., & Kalajdzic, A. (2019). Unpacking backlash: Individual and contextual moderators of bias against female professors. *Basic and Applied Social Psychology*, *41*(5), 305–325. https://doi.org/10.1080/01973533.2019.1652178

Fiske, S. (1993). Controlling other people: The impact of power on stereotyping. *American Psychologist*, *48*(6), 621–628. https://doi.org/10.1037/0003-066X.48.6.621

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Ford, D., Harkins, A., & Parnin, C. (2017). Someone like me: How does peer parity influence participation of women on stack overflow? *2017 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 239–243. https://doi.org/10.1109/VLHCC.2017.8103473

Galinsky, A. D., Gruenfeld, D. H., & Magee, J. C. (2003). From power to action. *Journal of Personality and Social Psychology*, *85*(3), 453–466. https://doi.org/10.1037/0022-3514.85.3.453

Gefen, D., & Ridings, C. M. (2005). If you spoke as she does, sir, instead of the way you do: A sociolinguistics perspective of gender differences in virtual communities. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, *36*(2), 78–92. https://doi.org/10.1145/1066149.1066156

Ginossar, T. (2008). Online Participation: A content analysis of differences in utilization of two online cancer communities by men and women, patients and family members. *Health Communication*, *23*(1), 1–12. https://doi.org/10.1080/10410230701697100

Glick, P., & Fiske, S. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*(3), 491–512.

Gruber, J., Mendle, J., Lindquist, K. A., Schmader, T., Clark, L. A., Bliss-Moreau, E., Akinola, M., Atlas, L., Barch, D. M., Barrett, L. F., Borelli, J. L., Brannon, T. N., Bunge, S. A., Campos, B., Cantlon, J., Carter, R., Carter-Sowell, A. R., Chen, S., Craske, M. G., … Williams, L. A. (2021). The future of women in psychological science. *Perspectives on Psychological Science*, 1745691620952789. https://doi.org/10.1177/1745691620952789

Guiller, J., & Durndell, A. (2006). 'I totally agree with you': Gender interactions in educational online discussion groups. *Journal of Computer Assisted Learning*, *22*(5), 368–381. https://doi.org/10.1111/j.1365-2729.2006.00184.x

Hall, J. A., Coats, E. J., & LeBeau, L. S. (2005). Nonverbal behavior and the vertical dimension of social relations: A meta-analysis. *Psychological Bulletin*, *131*(6), 898–924. https://doi.org/10.1037/0033-2909.131.6.898

He, J., Kang, S., & Lacetera, N. (2019). *Leaning In or Not Leaning Out? Opt-Out Choice Framing Attenuates Gender Differences in the Decision to Compete* (Working Paper No. 26484; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w26484

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior*, *32*, 113–135. https://doi.org/10.1016/j.riob.2012.11.003

Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks?:

The implied communality deficit. *Journal of Applied Psychology*, *92*(1), 81–92. https://doi.org/10.1037/0021-9010.92.1.81

Heilman, M., Wallen, A., Fuchs, D., & Tamkins, M. (2004). Penalties for success: Reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology*, *89*(3), 416–427.

Hinsley, A., Sutherland, W. J., & Johnston, A. (2017). Men ask more questions than women at a scientific conference. *PLOS ONE*, *12*(10), e0185534. https://doi.org/10.1371/journal.pone.0185534

Hofstra, B., Kulkarni, V. V., Galvez, S. M.-N., He, B., Jurafsky, D., & McFarland, D. A. (2020). The diversity–innovation paradox in science. *Proceedings of the National Academy of Sciences*, *117*(17), 9284–9291. https://doi.org/10.1073/pnas.1915378117

Hopkins-Doyle, A., Sutton, R., Douglas, K., & Calogero, R. (2019). Flattering to deceive: Why people misunderstand benevolent sexism. *Journal of Personality and Social Psychology*, *116*(2), 167–192. https://doi.org/10.1037/pspa0000135

Kaatz, A., & Carnes, M. (2014). Stuck in the out-group: Jennifer can't grow up, Jane's invisible, and Janet's over the hill. *Journal of Women's Health*, *23*(6), 481–484. https://doi.org/10.1089/jwh.2014.4766

Käfer, J., Betancourt, A., Villain, A. S., Fernandez, M., Vignal, C., Marais, G. A. B., & Tenaillon, M. I. (2018). Progress and prospects in gender visibility at SMBE Annual Meetings. *Genome Biology and Evolution*, *10*(3), 901–908. https://doi.org/10.1093/gbe/evy056

Kang, C. (2014, April 2). Google data-mines its approach to promoting women. *Washington Post*. https://www.washingtonpost.com/news/the-switch/wp/2014/04/02/google-data-mines-its-women-problem/

Kanter, R. M. (1977). *Men and women of the corporation*. Basic Books. https://agris.fao.org/agris-search/search.do?recordID=US201300554438

Keltner, D., Gruenfeld, D., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review*, *110*(2), 265–284.

Knobloch-Westerwick, S., Glynn, C. J., & Huge, M. (2013). The Matilda Effect in science communication: An experiment on gender bias in publication quality perceptions and collaboration interest. *Science Communication*, *35*(5), 603–625. https://doi.org/10.1177/1075547012472684

Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin*, *137*(4), 616–642. https://doi.org/10.1037/a0023557

Kray, L. J., & Thompson, L. (2004). Gender stereotypes and negotiation performance: An examination of theory and research. *Research in Organizational Behavior*, *26*, 103–182. https://doi.org/10.1016/S0191-3085(04)26004-X

Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Bibliometrics: Global gender disparities in science. *Nature News*, *504*(7479), 211. https://doi.org/10.1038/504211a

Lee, J. J., & Mccabe, J. M. (2021). Who speaks and who listens: Revisiting the chilly climate in college classrooms. *Gender & Society*, *35*(1), 32–60. https://doi.org/10.1177/0891243220977141

Leslie, S.-J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *347*(6219), 262–265.

https://doi.org/10.1126/science.1261375
Leventhal, H., Singer, R., & Jones, S. (1965). Effects of fear and specificity of recommendation upon attitudes and behavior. *Journal of Personality and Social Psychology*, *2*(1), 20–29. https://doi.org/10.1037/h0022089
Lewin, K. (1947). Frontiers in group dynamics: II. Channels of group life; social planning and action research. *Human Relations*, *1*(2), 143–153. https://doi.org/10.1177/001872674700100201
Lewin, K. (1952). Group decision and social change. In *Readings in Social Psychology*. Newcombe and Hartley (Eds.), Henry Holt.
Lips, H. M., & Lips, H. (1991). *Women Men And Power*. Mayfield Pub Co.
Llorens, A., Tzovara, A., Bellier, L., Bhaya-Grossman, I., Bidet-Caulet, A., Chang, W. K., Cross, Z. R., Dominguez-Faus, R., Flinker, A., Fonken, Y., Gorenstein, M. A., Holdgraf, C., Hoy, C. W., Ivanova, M. V., Jimenez, R. T., Jun, S., Kam, J. W. Y., Kidd, C., Marcelle, E., … Dronkers, N. F. (2021). Gender bias in academia: A lifetime problem that needs solutions. *Neuron*, *109*(13), 2047–2074. https://doi.org/10.1016/j.neuron.2021.06.002
Maranto, C. L., & Griffin, A. E. (2011). The antecedents of a 'chilly climate' for women faculty in higher education. *Human Relations*, *64*(2), 139–159. https://doi.org/10.1177/0018726710377932
Markus, H. R., & Kitayama, S. (2010). Cultures and selves: A cycle of mutual constitution. *Perspectives on Psychological Science*, *5*(4), 420–430. https://doi.org/10.1177/1745691610375557
Mast, M. S. (2002). Dominance as expressed and inferred through speaking time: A meta-analysis. *Human Communication Research*, *28*(3), 420–450. https://doi.org/10.1111/j.1468-2958.2002.tb00814.x
McCluney, C. L., & Rabelo, V. C. (2019). Conditions of visibility: An intersectional examination of Black women's belongingness and distinctiveness at work. *Journal of Vocational Behavior*, *113*, 143–152. https://doi.org/10.1016/j.jvb.2018.09.008
Milkman, K. L., Akinola, M., & Chugh, D. (2012). Temporal distance and discrimination: An audit study in academia. *Psychological Science*, *23*(7), 710–717. https://doi.org/10.1177/0956797611434539
Mo, P. K. H., Malik, S. H., & Coulson, N. S. (2009). Gender differences in computer-mediated communication: A systematic literature review of online health-related support groups. *Patient Education and Counseling*, *75*(1), 16–24. https://doi.org/10.1016/j.pec.2008.08.029
Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, *109*(41), 16474–16479. https://doi.org/10.1073/pnas.1211286109
Moss-Racusin, C. A., & Rudman, L. A. (2010). Disruptions in women's self-promotion: The backlash avoidance model. *Psychology of Women Quarterly*, *34*(2), 186–202. https://doi.org/10.1111/j.1471-6402.2010.01561.x
Murphy, M. C., Mejia, A. F., Mejia, J., Yan, X., Cheryan, S., Dasgupta, N., Destin, M., Fryberg, S. A., Garcia, J. A., Haines, E. L., Harackiewicz, J. M., Ledgerwood, A., Moss-Racusin, C. A., Park, L. E., Perry, S. P., Ratliff, K. A., Rattan, A., Sanchez, D. T., Savani, K., … Pestilli, F. (2020). Open science, communal culture, and women's participation in the movement to improve science. *Proceedings of the National Academy of Sciences*,

117(39), 24154–24164. https://doi.org/10.1073/pnas.1921320117

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science, 18*(10), 879–885.

Nittrouer, C. L., Hebl, M. R., Ashburn-Nardo, L., Trump-Steele, R. C. E., Lane, D. M., & Valian, V. (2018). Gender disparities in colloquium speakers at top universities. *Proceedings of the National Academy of Sciences*, *115*(1), 104–108. https://doi.org/10.1073/pnas.1708414115

Phelan, J. E., Moss-Racusin, C. A., & Rudman, L. A. (2008). Competent yet out in the cold: Shifting criteria for hiring reflect backlash toward agentic women. *Psychology of Women Quarterly, 32*(4), 406–413. https://doi.org/10.1111/j.1471-6402.2008.00454.x

Phelan, J., & Rudman, L. (2010). Reactions to ethnic deviance: The role of backlash in racial stereotype maintenance. *Journal of Personality and Social Psychology, 99*(2), 265–281. https://doi.org/10.1037/a0018304

Pritchard, J., Masters, K., Allen, J., Contenta, F., Huckvale, L., Wilkins, S., & Zocchi, A. (2014). Asking gender questions. *Astronomy & Geophysics, 55*(6), 6.8-6.12. https://doi.org/10.1093/astrogeo/atu245

Ridgeway, C. L., & Correll, S. J. (2004). Unpacking the gender system: A theoretical perspective on gender beliefs and social relations. *Gender & Society, 18*(4), 510–531. https://doi.org/10.1177/0891243204265269

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. Mcgraw-Hill.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Rudman, L. A. (1998). Self-promotion as a risk factor for women: The costs and benefits of counterstereotypical impression management. *Journal of Personality and Social Psychology, 74*(3), 629–645. https://doi.org/10.1037/0022-3514.74.3.629

Rudman, L. A., & Fairchild, K. (2004). Reactions to counterstereotypic behavior: The role of backlash in cultural stereotype maintenance. *Journal of Personality and Social Psychology*, *87*(2), 157–176. https://doi.org/10.1037/0022-3514.87.2.157

Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology*, *48*(1), 165–179. https://doi.org/10.1016/j.jesp.2011.10.008

Rudman, L., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology*, *77*(5), 1004–1010.

Schmader, T. (2010). Stereotype threat deconstructed. *Current Directions in Psychological Science*, *19*(1), 14–18. https://doi.org/10.1177/0963721409359292

Schmidt, S. J., & Davenport, J. R. A. (2017). Who asks questions at astronomy meetings? *Nature Astronomy*, *1*(6), 1–2. https://doi.org/10.1038/s41550-017-0153

Schroeder, J., Dugdale, H. L., Radersma, R., Hinsch, M., Buehler, D. M., Saul, J., Porter, L., Liker, A., Cauwer, I. D., Johnson, P. J., Santure, A. W., Griffin, A. S., Bolund, E., Ross, L., Webb, T. J., Feulner, P. G. D., Winney, I., Szulkin, M., Komdeur, J., … Horrocks, N. P. C. (2013). Fewer invited talks by women in evolutionary biology symposia. *Journal of Evolutionary Biology*, *26*(9), 2063–2069. https://doi.org/10.1111/jeb.12198

Shen, H. (2013). Inequality quantified: Mind the gender gap. *Nature News*, *495*(7439), 22. https://doi.org/10.1038/495022a

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, *26*(5), 559–569. https://doi.org/10.1177/0956797614567341

Simpson, R., & Lewis, P. (2005). An investigation of silence and a scrutiny of transparency: Re-examining gender in organization literature through the concepts of voice and visibility. *Human Relations*, *58*(10), 1253–1275. https://doi.org/10.1177/0018726705058940

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613–629.

Sullivan, J., Moss-Racusin, C., Lopez, M., & Williams, K. (2018). Backlash against gender stereotype-violating preschool children. *PLOS ONE*, *13*(4), e0195503. https://doi.org/10.1371/journal.pone.0195503

Tannen, D. (1993). *Gender and Conversational Interaction*. Oxford University Press.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, *36*(7), 778–793. https://doi.org/10.1037/0022-3514.36.7.778

Telis, N., Glassberg, E. C., Pritchard, J. K., & Gunter, C. (2019). Public discussion affects question asking at academic conferences. *The American Journal of Human Genetics*, *105*(1), 189–197. https://doi.org/10.1016/j.ajhg.2019.06.004

Vásárhelyi, O., Zakhlebin, I., Milojević, S., & Horvát, E.-Á. (2021). Gender inequities in the online dissemination of scholars' work. *Proceedings of the National Academy of Sciences*, *118*(39). https://doi.org/10.1073/pnas.2102945118

Vasilescu, B., Capiluppi, A., & Serebrenik, A. (2012). Gender, representation and online participation: A quantitative study of StackOverflow. *2012 International Conference on Social Informatics*, 332–338. https://doi.org/10.1109/SocialInformatics.2012.81

Vosgerau, J., Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2019). 99% impossible: A valid, or falsifiable, internal meta-analysis. *Journal of Experimental Psychology. General*, *148*(9), 1628–1639. https://doi.org/10.1037/xge0000663

Walker, R. L., Sykes, L., Hemmelgarn, B. R., & Quan, H. (2010). Authors' opinions on publication in relation to annual performance assessment. *BMC Medical Education*, *10*(1), 21. https://doi.org/10.1186/1472-6920-10-21

Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology*, *92*(1), 82–96. https://doi.org/10.1037/0022-3514.92.1.82

Wang, Y. S., Lee, C. J., West, J. D., Bergstrom, C. T., & Erosheva, E. A. (2019). Gender-based homophily in collaborations across a heterogeneous scholarly landscape. *ArXiv:1909.01284 [Stat]*. http://arxiv.org/abs/1909.01284

Wood, W., & Eagly, A. H. (2012). Chapter two—Biosocial construction of sex differences and similarities in behavior. In J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (Vol. 46, pp. 55–123). Academic Press. https://doi.org/10.1016/B978-0-12-394281-4.00002-7

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, *330*(6004), 686–688. https://doi.org/10.1126/science.1193147

Appendix 1: Q&A Session Behavioral Codebook
Summer 2020
Shoshana Jarvis, Charlie Ebersole, Christine Nguyen, Leila Zhu
==================================

General Guidelines for Timing:

- Round measurements to the nearest whole second
- Start the time when the person begins speaking and end the time when person stops speaking (don't count dead air between individuals)
- For interruptions: if it's a clarification to help the speaker, it counts in the initial remark section. If it's a second question, it's a follow up question and counted in total time.

1) Time of Question (Initial Remark)

Definition: The amount of time spent in the initial question or comment the question asker poses to the speaker.

Integer, in seconds

2) Time Spent Introducing Themselves

Definition: If the question asker introduces themselves, the amount of time they spent on their introduction. This will be a subset of the initial remark. It should refer to information about the speaker. Comments like the person's name or their field of study count as introductions.

Integer, in seconds

3) Time of Speaker Response (Total)

Definition: The amount of time the speaker spends responding to a question/comment. This is total time and should encompass responses to any follow up questions/comments.

Integer, in seconds

4) Time of Question Asker (Total)

Definition: The total amount of time that the question asker spends speaking. This includes any follow up comments or questions.

Integer, in seconds

5) Gender Presentation of Question Asker

Definition: With your best guess, does the question asker seem male, female, or gender non-binary

1 = male
2 = female
3 = non-binary

6) Number of Questions (Initial Remark)

Definition: The number of questions asked in the initial remark.

Count

7) Number of Comments (Initial Remark)

Definition: The number of comments posed in the initial remark.

Count

Examples: These should be distinct from setting up context for a question. As an example saying something like "In economics, we release working papers long before research is published in a journal. How do you think that process would work in psychology?" is just a question, not a comment. The initial declarative statement is just there to explain the question.

8) Number of Comment/Questions (Initial Remark)

Definition: The number of comment/questions in the initial remark. A comment question is primarily a comment that the question asker tries to make a question and/or adds a request for reactions at the end.

Count

Examples: Question askers will sometimes try to turn a comment into a question by adding phrases like "what are your reactions to that?", "I'd love your thoughts on that", or by changing their vocal inflection at the end of a statement.

9) Number of Question/Comments (Initial Remark)

Definition: The number of question/comments in the initial remark. A question comment is when a question asker poses a question and then follows it up with their own commentary before the speaker responds.

Count

Examples: Common question/comments are when a question asker poses a question that they then provide their own answer for before giving the speaker time to respond.

10) Number of Follow up Questions

Definition: The number of questions the question asker asks after the speaker responds to the initial remark.

Count

11) Number of Follow up Comments

Definition: The number of comments the question asker says after the speaker responds to the initial remark.

Count

12) Number of Follow up Comment/Questions

Definition: The number of comment/questions the question asker says after the speaker responds to the initial remark.

Count

13) Number of Follow up Question/Comments

Definition: The number of question/comments the question asker says after the speaker responds to the initial remark.

Count

14) Thank You for Talk

Definition: Did the question asker thank the speaker for their presentation/talk/time.

0 = no
1 = yes

15) Thank You for Response

Definition: Did the question asker thank the speaker for their response.

0 = no
1 = yes

16) Initial Positive Comment

Definition: Does the question asker, near the beginning of their question and/or comment, make a positive comment about the talk.

0 = no
1 = yes

Examples: "Interesting talk", "Fascinating talk", "that was a great talk"

17) Talking over speaker

Definition: How often does the question asker talk over the speaker. These are intentional interruptions of the speaker (rather than accidental interruptions). These instances do not need to include both people speaking at the same time. If it is clear that the speaker had more to say but the asker jumped in while they had paused (e.g., to breathe), this is an interruption. Brief interjections (e.g., "totally agree") by the asker that don't disrupt the flow of the speaker/their comments do not count as interruptions. Finally, interruptions can be positive in valence, they just need derail or disrupt the speaker (e.g., interrupting to agree with the speaker and then giving extended comments).

Count

Examples: We are interested in intentional interruptions. For these, the question asker should being speaking while the speaker is still talking and complete at least one thought. By contrast, accidental interruptions will likely be followed by some communication between the asker and speaker to determine who should continue speaking (e.g., “you got first”, “sorry, finish your thought”).

18) Reference own research

Definition: Does the question asker reference their own research at some point. This does not include information that is offered during an introduction. Mentioning someone else’s research doesn’t count; it has to be their own work. Likewise, talking about one’s field, as a whole, doesn’t count (e.g., “in my field of psychology, we…).

0 = no
1 = yes

Examples: “In my work, we find…”, “I have a paper that shows…”, “I’ve been looking at this for 20 years and…”. Saying in an intro, “I’m Charlie and I study metascience” doesn’t count. However saying, “Hi, I’m Charlie, and I’m curious about your work because in my research in metascience…” does.

19) Valence of the referencing own work

Definition: If the question asker referenced their own work, did they reference it in a positive, neutral, or negative way.

-1 = negative
0 = neutral
1 = positive
if they don’t reference their own work, leave this blank

Examples: Positive references are ones where the asker seems to be bragging about or promoting their own work or themselves. Neutral references are those that simply state the asker’s research/area of research (for instance as context). Negative references could include stating that the asker has made a similar mistake to what was addressed in a talk (e.g., “you talked about how publication bias is bad, in my research, we totally biased what we published”).

20) Challenges speaker

Definition: Does the question asker, at any point, challenge the speaker in terms of a claim they made or their expertise. If there’s an implication that a researcher is poor at research or is manipulating their results, that counts as a challenge. Suggestions do not count as challenges. Challenges should indicate that the question asker believes the speaker did something improper or wrong.

0 = no
1 = yes

Examples: More direct confrontations could include things like “I think that’s wrong”, “I’ve seen research that says the opposite.” More indirect confrontations could include phrases like “I’m concerned about this part of your talk…” or “well in economics, we’ve studied this for 20 years and…”. Challenges can also be positive as in “I think you should have accounted for demographics in your model” but said in a nice way. “Have you thought about this?” is not a challenge, just a suggestion.

Appendix 2: Q&A Questions Qualitative Coding Codebook
Summer 2020
Shoshana Jarvis, Charlie Ebersole, Christine Nguyen, Leila Zhu

==================================

1) Wanting to make space for others

Definition: Does the attendee indicate that they altered their question asking behavior in order to provide more space and/or speaking time for other attendees.

0 = no mention
1 = yes

Examples:

0: R_2mDL04SDrVJ7Sg1
Typically, it is because I assume my question is general knowledge and is something I just don't know yet because I am new to the field.

1: R_31heCEaIpmVbwcI
Often to let students ask.

2) Logistical/Practical Constraints

Definition: Does the attendee indicate that they altered their question asking behavior due to logistical and/or practical constraints, such as being seated far from the microphone or being unable/unwilling to try to get the attention of the moderators.

0 = no mention
1 = yes

Examples:

0: R_2mDL04SDrVJ7Sg1
Typically, it is because I assume my question is general knowledge and is something I just don't know yet because I am new to the field.

1: R_3r2EsQ8S26GpWd9
Usually because I don't have time nor opportunity to approach the speaker.

3) Insecurity/Anxiety

Definition: Does the attendee indicate that they altered their question asking behavior because of personal insecurities or anxieties, such as being afraid to speak in front of the room or worrying about sounding incompetent.

0 = no mention
1 = yes

Examples:

0: R_3r2EsQ8S26GpWd9
Usually because I don't have time nor opportunity to approach the speaker.

1: R_3rPxditL9yjQkxf
I always have trouble initiating a conversation with people I'm not familiar with

4) Prefer 1-on-1

Definition: Does the attendee indicate that they chose not to engage in the Q&A session because they prefer to talk to speakers 1-on-1.

0 = no mention
1 = yes

Examples:

0: R_3r2EsQ8S26GpWd9
Usually because I don't have time nor opportunity to approach the speaker.

1: R_Aj1NZSRws3dCXSN
Sometimes I hold back on a question because I find it more comfortable to approach the person afterward and have more of a conversation than to ask a question in front of a large group. I sometimes ask questions during the QandA, but I never feel completely comfortable doing so.

5) Fear of backlash

Definition: Does the attendee indicate that they altered their question asking behavior they were afraid of negative repercussions for themselves if they were to ask a question.

0 = no mention
1 = yes

Examples:

0: R_2YPJNxvc7GvghQJ
Other people have questions and are ahead of me in the queue.

1: R_3nf7uBCyFCvydOo
Limited time. Potential negative consequences for publishing and employment.

6) Negative audience attitudes

Definition: Does the attendee indicate that they altered their question asking behavior because the audience is hostile to questions.

0 = no mention
1 = yes

Examples:

0: R_2YPJNxvc7GvghQJ
Other people have questions and are ahead of me in the queue.

1: R_rqZAsTVSB4LLzvH
I think question and answer frames the interaction as interacting with an authority. While the presenter is the authority on their work, I think they are also not An Authority, and I think the two are easily conflated (by the presenter and the audience) and the situation becomes quite threatening to approach. I'd rather the format of presentations be reframed.

7a) Identity factor - race

Definition: Does the attendee indicate that they altered their question asking behavior because of their race.

0 = no mention
1 = yes

Examples:

0: R_2q7tJ9PGYn3wGZX
First of all my English - I'm not a native English speaker. Second, I'm afraid to look dumb by asking dumb questions.

1: R_3DdKzQD1Of2aZKX
Racial, gender, rank, and age bias also lead some to dismiss the very crucial ideas brought forward by others, especially if the ideas challenge the issues of psych research's WEIRD focuses.

7b) Identity factor - international status

Definition: Does the attendee indicate that they altered their question asking behavior because of their international status.

0 = no mention
1 = yes

Examples:

0: R_UKGic7UcPDasF4R
A female grad student asking questions isn't always viewed the most favorably.

1: R_2q7tJ9PGYn3wGZX
First of all my English - I'm not a native English speaker. Second, I'm afraid to look dumb by asking dumb questions.

7c) Identity factor - gender

Definition: Does the attendee indicate that they altered their question asking behavior because of their gender.

0 = no mention
1 = yes

Examples:

0: R_3EzCci3F02MkcV8
Early career researcher and do not want to come off as unknowledgable.

1: R_UKGic7UcPDasF4R
A female grad student asking questions isn't always viewed the most favorably.

7d) Identity factor - rank

Definition: Does the attendee indicate that they altered their question asking behavior because of a status-level threat -- i.e. undergrads, lab managers, grad students, or even assistant professors who indicate that their rank is not high enough to say a question.

0 = no mention
1 = yes

Examples:

0: R_2q7tJ9PGYn3wGZX
First of all my English - I'm not a native English speaker. Second, I'm afraid to look dumb by asking dumb questions.

1: R_UKGic7UcPDasF4R
A female grad student asking questions isn't always viewed the most favorably.

8) Irrelevance to the larger group

Definition: Does the attendee indicate that they altered their question asking behavior because their question is potentially irrelevant to the larger group. (To the extent it overlaps with other categories (one-on-one, general anxiety, etc) it can count for both)

0 = no mention
1 = yes

Examples:

0: R_V4sR8LCdCBzn0eR
Feeling like it's not a good question or I might have missed it during the presentation.

1: R_3iW6iMMeoAodSzn
Because sometimes they're relatively personal questions that might not apply to the general public, and therefore, it would be easier to ask them one-on-one instead.

9) Not wanting to challenge

Definition: Does the attendee indicate that they altered their question asking behavior because they did not want to seem like they were challenging the speaker?

0 = no mention
1 = yes

Examples:

0: R_puaNn1TpAZviK1X
Fear of sounding silly or unintelligent

1: R_Q0xtrmYeLjo7Ed3
Not sure if they are important enough. I do not want to appear as though I am critiquing their work. I sometimes just read the paper when it is published and often, my questions are answered there.