Public Discussion Affects Question Asking at Academic Conferences

Natalie Telis,[1,2,8,*] Emily C. Glassberg,[3,8] Jonathan K. Pritchard,[2,3,4] and Chris Gunter[5,6,7]

Women are under-represented in science, technology, engineering, and mathematics (STEM). Despite the recent emphasis on diversity in STEM, our understanding of what drives differences between women and men scientists remains limited. This, in turn, limits our ability to intervene to level the playing field. To quantify the representation and participation of women and men at academic meetings in human genetics, we developed high-throughput and crowd-sourced approaches focused on question-asking behavior. Question asking is one voluntary and self-initiated scientific activity we can measure. Here we report that women ask fewer questions than expected regardless of their representation in talk audiences. We present evidence that external barriers affect the representation of women in STEM. However, differences in question-asking behavior suggest that internal factors also impact women's participation. We then examine the effects of specific interventions and show that wide public discussion of the relative under-participation of women in question-and-answer sessions alters question-asking behavior. We suggest that engaging the community in such projects promotes visibility of diversity issues at academic meetings and allows for efficient data collection that can be used to further explore and understand differences in conference participation.

Introduction

We believe that science should be a place where all have an equal opportunity to succeed. To date, however, opportunities to participate and succeed in the scientific enterprise vary by race, gender, and socio-economic status. In this study, we focus on differences between women and men scientists (using a binary gender classification, recognizing that this does not fully capture the spectrum of gender identity).

Disparities between women and men are well characterized throughout academic publishing. Across disciplines, women publish and are cited less than men.[1–3] One journal's own data recently showed that women appear to be disadvantaged at every step of the review process.[4] Women are less likely to be first authors and are more likely to be acknowledged as performing, rather than designing, published work.[1,5–7] These trends are concerning for women scientists, as citations and first-author publications are used to measure academic merit.

Such disparities persist outside of the literature; women are under-represented as speakers at academic meetings in fields as diverse as microbiology,[8] virology,[9] evolutionary biology,[10,11] anthropology/primatology,[12] and social psychology.[13] Conference presentations provide an opportunity to garner feedback, visibility, and prestige for oneself and one's work. It is therefore likely that this trend hinders the careers of women scientists.

The need to address this glaring under-representation seems straightforward; however, effective solutions have proved elusive. Subtle factors such as unconscious bias, differential description of research in applications, and self-selection may all impact the gender ratio of conference speakers.[10,14] Such factors may also influence the overall under-representation of women in the sciences.[15]

Further, achieving equal representation is not sufficient to ensure equal opportunities for success for women and men. In addition to differences in representation, a growing body of evidence suggests that there are large differences in participation by gender. In particular, studies from a wide range of fields suggest that women under-participate in question-and-answer sessions following talks.[16–22]

Understanding differences in question asking by gender is especially interesting for several reasons. First, question asking is voluntary and self-initiated. Gendered differences in question asking are therefore unlikely to arise solely from conscious bias or overt discrimination. They may instead arise from more subtle forces (e.g., self-selection) that are often more challenging to measure and address.

Second, asking questions at conferences may help researchers advance their scientific careers. In STEM classrooms, question asking has been identified as a mechanism to encourage higher-order cognitive thinking and improve retention.[23–25] Extrapolating to adulthood, formulating questions may help conference attendees incorporate new information into their own research. Asking questions may also be a way to garner esteem among colleagues.

Third, even if asking questions does not directly impact one's likelihood of academic success, question asking may be correlated with traits that do. For instance, men may be

[1]Biomedical Informatics Program, Stanford University, Stanford, CA 94305, USA; [2]Department of Genetics, Stanford University, Stanford, CA 94305, USA; [3]Department of Biology, Stanford University, Stanford, CA 94305, USA; [4]Howard Hughes Medical Institute, Janelia Farms, Ashburn, VA 20815, USA; [5]Department of Pediatrics, Emory University School of Medicine, Atlanta, GA 30322, USA; [6]Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322, USA; [7]Marcus Autism Center, Children's Healthcare of Atlanta, Atlanta, GA 30329, USA
[8]These authors contributed equally to this work
*Correspondence: natalie.telis@gmail.com
https://doi.org/10.1016/j.ajhg.2019.06.004.

more comfortable with self-promotion,[26,27] which may lead them to ask more questions and to apply for more grants, etc. Understanding differences in participation in question-and-answer sessions may provide insight into subtle drivers of the continuing gender disparity in the sciences.

To characterize disparities in both the representation and participation of women and men scientists, we quantified gender ratios of poster presenters, oral presenters, and question askers at human genetics meetings. We identified several behavioral differences in question-and-answer sessions that hint at subtle and unconscious drivers of the under-participation of women. We then intervened by stimulating wide public discussion of these trends. In response, we observed marked shifts in question-asking behavior. Finally, we developed a novel crowd-sourced method for recording participation in question-and-answer sessions. This approach lays the groundwork for future high-throughput data collection and exploration.

Material and Methods

Gender Inference for Demographic Analysis

A summary of data collected from annual meetings of the American Society of Human Genetics (ASHG) and Biology of Genomes (BoG) is shown in Table S1. Sessions from which question data were recorded at ASHG, as well as their associated topics, are listed in Table S2.

Publicly available abstract booklets and online presenter registries for the American Society of Human Genetics (ASHG) 2014–2017 annual meetings were automatically downloaded and parsed using python packages requests and urllib2. For each of 13,138 unique presentations, we collected first author, presentation session, and presentation type: poster, platform talk (concurrent oral presentation by application), plenary talk (conference-wide oral presentation by application), and invited talk (presentation by specific invitation by a session organizer). For poster presenters, we used presentation session as subfield membership. For oral presentations, we used presentation session as a proxy for subfield membership; we manually matched each oral presentation session to a corresponding poster session subfield.

We identified 9,271 unique presenters between 2014 and 2017. To infer gender for each of these presenters, we applied genderizeR, an extension of the genderize.io API[28] to presenter first names. We then captured rare US names and disambiguated some gender ambiguous names using a database of US Census data containing 216,286 distinct names from 79 countries in 89 languages, accessed on 1/26/18 through UTF-8 or Latin encodings (per the procedure from Larivière et al.[7]).

We validated our gender inference approach with aggregated, self-reported gender of ASHG attendees (2015–2016) provided by ASHG staff. The proportions of men and women inferred in each subfield were consistent with the proportions of self-reported men and women.

Question-and-Answer Data Collection

Two raters, E.C.G. and N.T., recorded questions in person at all sessions attended during the 2014–2016 ASHG annual meetings and the 2015–2018 BoG annual meetings. Sessions attended were heavily weighted toward our research interests, statistical and computational genetics. For each question, the speaker and session were recorded, as were the questioner gender, whether the questioner was a session moderator, and whether the question included any follow-up questions.

Crowd-Sourced Data Collection

In order to understand the scope of the effect, we believed it important to record data conference-wide, rather than just in sessions our three raters could attend. To this end, we hosted an online portal on Amazon Web Services for login-less entry of questions during the ASHG 2017 annual meeting (October 17th—October 21st). This online portal communicated with a secure database that included speaker and conference meta data. We enabled conference attendees to enter:

Exact speaker name (text field)
Search-and-select speaker using the database of ASHG presenters
Questioner moderator status (yes, no, or NA)
Questioner gender (man, woman, other, or NA)
These categories are a gross simplification of gender, and all records are based on raters' perceptions of questioner gender. Future studies with much larger sample sizes could and should explicitly include broader, and self-reported gender identities.
Follow-up question status (yes, no, or NA)
Was this a second question from the same questioner?
Questioner sentiment - challenge (yes, no, or NA)
Was the questioner challenging the speaker? Challenges might include the words "challenge," "justify," or "artifact."
Questioner sentiment - compliment (yes, no, or NA)
Did the question include a compliment? Compliments include "nice talk," "beautiful work," etc.

To ensure that our question data were not affected by false automated entry, we employed a reCAPTCHA human identification module. We also recorded the IP addresses of unique devices participating in question recording, as well as the time of each record, to enable exclusion of systematically improper entries. Finally, incomplete entries, or entries not tied to an identified speaker in the database, were excluded.

Crowd-Sourced Question Analysis

In our crowd-sourced data collection, raters recorded questions at their leisure. Some raters recorded only part of a question-and-answer session or skipped a question in the middle of a session. As a result, the 65% of talks with more than two raters often contained nonidentical question sequences. Previous approaches selecting the longest string of questions and determining it as canonical are likely to fail in any case where there is rater disagreement and do not take full advantage of the crowd-sourced data approach. Moreover, combining identical sequences of questions without thought to placement in time may in fact mask rater disagreement differences. We therefore developed an algorithm to reconstitute question-and-answer sessions from the recordings of disparate raters. Our algorithm is designed to deduplicate questions recorded by multiple raters, collapsing them into one "centroid" with multiple pieces of supporting near-contemporaneous evidence. This approach allows us to integrate data across

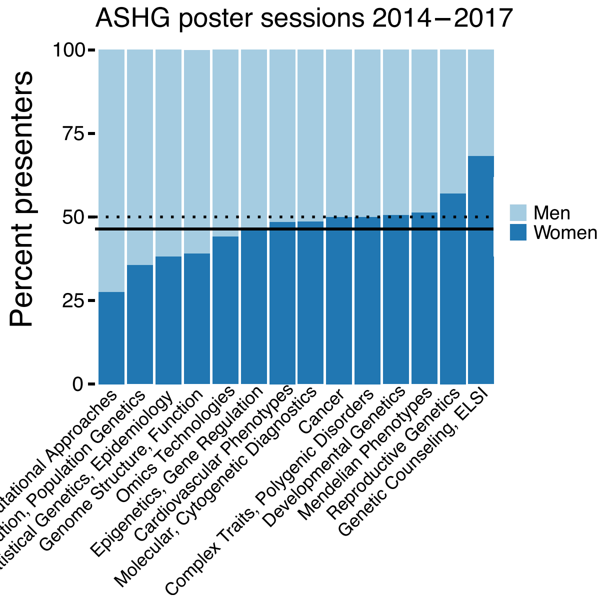

Figure 1. Representation of Women in Genetics Varies by Subfield

Data are 11,351 abstracts of poster presentations from the 2014–2017 annual meetings of the American Society of Human Genetics. Abstract books were mined for presenter names (first author) and subfields. Presenter gender inferred via genderizeR (see Material and Methods and Wais[28]). The percent of women poster presenters is in dark blue, men in light blue. Solid line marks the overall representation of women across subfields (45%), dotted marks equal representation (50%). Variance in male/female ratio across sessions is highly significant ($p < 0.001$ by multinomial permutation test).

multiple incomplete ratings and takes into account the timing of distinct records for individual talks.

Algorithm Procedure

Given an initialization set of questions $\{q_1, \ldots q_n\}$, we iterate over this procedure:

1. For each question q_i;
2. Examine each unique record from the crowd-sourced database
3. If this piece of evidence e_j meets all of the following requirements, consider it as supporting evidence for q_i:

e_j and the time of q_i are within some threshold t of one-another;
q_i has not already been supported by the unique submitter who recorded e_j;
q_i and e_j record the same gender

4. If e_j appears to be evidence for q_i, update the body of evidence for the question:

Add e_j to the end of the list of pieces of evidence supporting it, $\{e_{i,1}, \ldots e_{i,n}\}$
Set the time of record for q_i to be the mean of all the pieces of evidence supporting it

5. When no further evidence can be added to q_i, either continue to the next question, or choose a record from the remaining set of records to initialize a novel question, adding to the set above, $\{q_1, \ldots q_{n+1}\}$.

We continue this procedure until no records remain unassigned as evidence for a given speaker.

Filtering

We identify all questions with centroids that are extremely close to one another after filtering. From the 1,014 original records in the crowd-sourced data, we initially derive 656 unique question centroids. Of these, we identify only 36 that record the same gender and are within 10 s of one another; by manual examination all of these records appear so spuriously close that we believe they represent the same question and collapse them conservatively to one centroid. We identify only one instance of a mis-recorded gender (nearly identical centroid location, different gender recorded). After this filtration we are left with 638 total question records. 220 of them, or 34.5%, are from individuals identified as women. This is slightly lower than the proportion of women reported in the smaller, single-rater subset (38.5%) but is barely significantly different (binomial p value $p = 0.042$).

Results

To explore the role of gender at conferences in genetics, we collected and analyzed data from 4 years of annual meetings of the American Society of Human Genetics (ASHG). ASHG is a large and diverse academic and clinical society; its nearly 8,000 members include an array of genetics professionals such as experimental and computational researchers, physicians, ethicists, and genetic counselors. ASHG's annual meetings, therefore, provide an excellent forum in which to study the representation and participation of women across subfields of genetics. We also include data from 4 years of the Biology of Genomes (BoG) meeting. Unlike ASHG, which is a large conference with many concurrent sessions, BoG is a more concentrated meeting; all talks take place in a single auditorium. Analyzing both meetings allows us to understand the robustness of our results to meeting type and setting. A summary of the data collected from both meetings is shown in Table S1.

Participation of Women in Poster Sessions Varies by Subfield

We first characterized the demographics of ASHG at a granular subfield level to define a measure of representation. We mined data for 13,138 abstracts from the 2014–2017 annual meetings from publicly available abstract books. For each abstract, we used the established genderizeR algorithm[28] and census data to infer presenter (first-author) gender from first names (see Material and Methods). We also automated the extraction of the subfield of each abstract, measured as the session in which it was presented. The proportions of men and women presenting posters in each subfield are shown in Figure 1 (number of poster presentations in each subfield shown in Figure S1). We then corroborated our estimated proportions of men and women per subfield with data on authors' self-reported gender supplied by the ASHG staff (see Figure S3).

Some fields, like bioinformatics and computational approaches and genome structure and function, were significantly depleted for women relative to conference

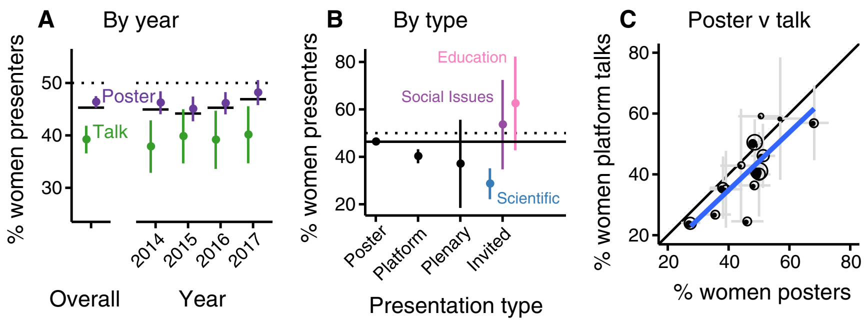

Figure 2. Women Are Under-represented in Oral Presentations
Data are abstracts from 1,787 talks and 11,351 poster presentations from the 2014–2017 annual meetings of the American Society of Human Genetics.
(A) Percent of women presenters by year. Black lines mark the overall proportion of women presenters for each year. Dotted line marks equal representation (50% women). Percent of speakers who are women (including invited, featured plenary, and platform talks) are shown in green, poster presenters in purple. Vertical lines mark ± 2 binomial standard errors.
(B) Percent of presenters who are women by presentation type across all four annual meetings. Presentations granted by application (poster, platform, and plenary) shown in black, those granted by invitation (invited scientific, social issues, and education) in blue, purple, and pink, respectively. Vertical lines mark ± 2 binomial standard errors. Dotted linemarks equal representation (50%), solid line marks overall representation of women presenters across four annual meetings (45%).
(C) Percent women poster presenters (x axis), compared with the percent of women presenting platform talks. Presentations were assigned to a subfield based on the subfield to which the corresponding abstract was submitted. Point size indicates the total number of presenters (poster and talk) in a subfield. Grey lines show ± 2 binomial standard errors. Black line indicates equal representation of women in posters and talks (y = x), blue line indicates the best-fit line of the true relationship between poster and platform presentations ($y = -0.03 + 0.95x$).

attendees overall (27% and 39% of women poster presenters, respectively; overall attendees 45% women, $p = 5.0 \times 10^{-30}$, $p = 8.6 \times 10^{-3}$ by binomial tests). Others, such as genetic counseling, ELSI, education, and health services research, and prenatal, perinatal, and reproductive genetics, were enriched for women (67% women, $p = 2.5 \times 10^{-19}$, 57% women; $p = 1.2 \times 10^{-3}$, respectively).

Overall, there was significantly greater variance in the proportion of women presenting posters across subfields than would be expected based on total conference attendance ($p < 0.001$ by multinomial permutation test). This high variance in gender ratio across subfields suggests that coarse-grained diversity metrics, such as the proportion of women among the 8,000+ conference attendees at ASHG, may fail to capture important features of representation in our field.

Women Are Under-represented in Oral Presentations

Regardless of representation within a subfield, women persistently presented fewer talks than expected at ASHG. In 2014–2017, women presented 39% of 1,367 talks for which we could confidently infer speaker gender (women are 45% of attendees, $p < 2.2 \times 10^{-16}$ by binomial test; Figure 2A). This effect is consistent across subfields; women consistently presented fewer talks than expected based on their participation in poster sessions of the corresponding subfield. The median difference between the percentage of women presenting posters and women presenting talks is above 10% (see Figure 2C).

One might hypothesize that this relative under-participation of women in talks is attributable to reviewer bias. However, the depletion of women presenters of platform talks existed in all 4 years studied. This is notable because, starting in the second year studied (2015), the ASHG Program Committee implemented a double-blind application process in which reviewers were unaware of the names (and genders) of abstract authors. This blinding did not have a significant effect on the number of talks versus posters awarded to women authors.

The depletion of talks from women may therefore result from more nuanced factors such as self-selection (for instance, women may be more likely to select the "Poster only" option upon abstract submission, which may decrease their chances of receiving a talk) or dialectical differences in how women and men present their findings in abstracts. Understanding the influence of such internal factors is important for developing interventions to increase the participation and success of women in STEM; removing external barriers such as potentially biased review processes may be insufficient to achieve parity between men and women scientists.

Evidence for the impact of internal factors does not, however, preclude effects of external biases. At ASHG 2014–2017, women presented a significantly lower proportion of invited scientific talks, which were awarded in an explicitly unblinded fashion, than platform talks, which, after 2015, were awarded by double-blind application (Figure 2B). This suggests that that the invitation process, which differs in several key ways from presentation by application (including the unblinded explicit invitation of specific speakers), adversely impacted the participation of women. The under-representation of women speakers, however, does not hold for the invited talks on education and social issues. In these talks, women were overrepresented relative to all conference attendees (Figure 2B).

Women Ask Fewer Questions than Men Regardless of Representation

Although presentation captures aspects of both women's representation and participation at ASHG, we hoped to extend our analysis to broader behavioral metrics. We categorize gendered differences in behavior with a focus on

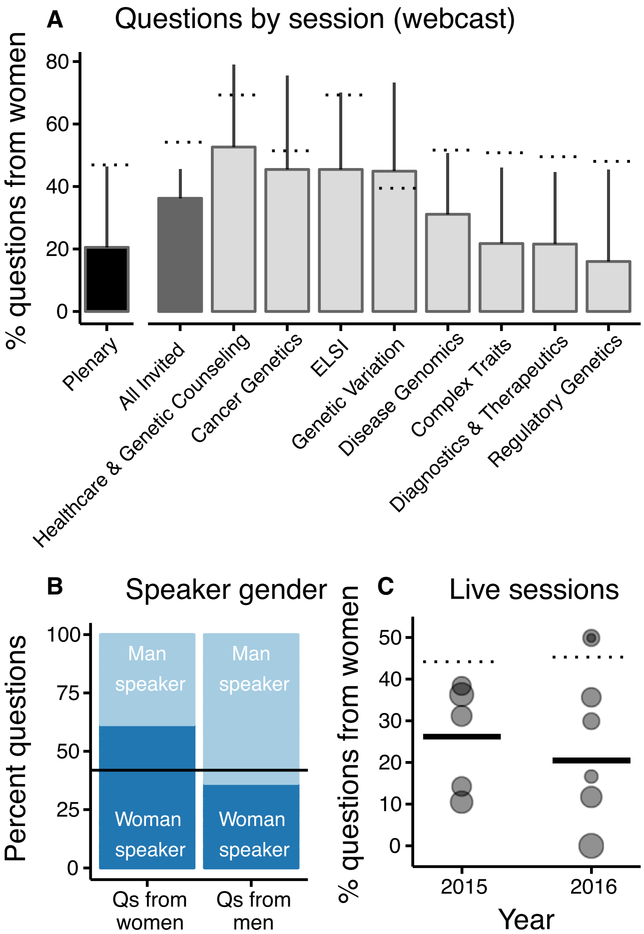

Figure 3. Women Are Under-represented in Question-and-Answer Sessions

(A) Percent of questions asked by women by talk category (webcast). Data are 146 questions recorded from webcasts of 160 invited talks from the 2014–2016 ASHG annual meetings. Plenary talks, which are attended by the entire meeting, are shown in black. The proportion of questions asked by women across all invited talks is shown in dark gray, and the proportions in specific subfields in light gray. Dotted lines show the expected proportion of questions from women, based on the proportion of women poster presenters in the corresponding subfield. Talks are manually assigned to subfields (i.e., poster sessions) based on their titles. For plenary talks, the audience is estimated as the overall proportion of women attending the meeting.

(B) Effect of speaker gender on question-and-answer sessions. Filled bars show the proportion of questions from women and men, respectively (x axis) addressed to women (dark blue) and men (light blue) speakers. Black bar indicates the overall proportion of women speakers (and, therefore, the expected proportion of questions addressed to women). Data from invited and plenary talks from 2014–2016 ASHG annual meetings, as in (A).

(C) Percent of questions asked by women (live). Data are questions recorded live during sessions attended by N.T. and E.C.G. at the 2015 and 2016 ASHG annual meetings. Each point shows the proportion of questions asked by women in a single session. Point size represents the number of talks in each subfield. Solid lines show the proportion of questions asked by women across subfields, dotted lines show the proportion of women presenters (and, therefore, the expected proportion of questions from women) for each year.

question asking during question-and-answer (Q&A) sessions following oral presentations.

To test for differences in question asking between men and women, we recorded questions asked following the invited and plenary talks from the 2014–2016 ASHG annual meetings (available via webcast). We estimated the proportion of women in the audience of each invited session as the proportion of women poster presenters in the corresponding subfield (invited sessions were matched manually to subfields). Plenary talks were assumed to capture the entire meeting audience; the proportion of women in the audience was estimated as the overall proportion of women presenters (poster and oral) at the conference. Finally, for sessions in each subfield, we tested the proportion of questions asked by women against the estimated gender ratio of the audience.

We find that women asked fewer questions than expected based on estimated audience proportions (women asked 35% of 463 recorded questions, average audience across invited and plenary talks was estimated to be 53% women, $p < 4 \times 10^{-15}$ by binomial test). Interestingly, this was true even in women-majority subfields such as genetic counseling and ELSI; women were expected to comprise 67% of ELSI audiences, yet only 45% of questions after ELSI talks were asked by women (Figure 3A). This argues against a hypothesis that women are discouraged from asking questions because, as the minority, they consider themselves outsiders. The fact that genetic counseling and ELSI have historically been dominated by women suggests that seniority alone does not drive female question-asking behavior.

Given that differences in question asking between men and women seem unlikely to be driven by majority/minority effects or by differences in seniority, we explored other factors that affected the proportion of questions from women. In particular, we noted that women asked a disproportionate number of their questions to women speakers. Conversely, men asked a disproportionate number of their questions to men (58% of 161 questions from women went to women speakers, 63% of 302 questions from men went to men, $p < 3 \times 10^{-5}$, $p < 0.08$, respectively, by binomial tests against the proportions of women and men speakers). This relationship between speaker gender and question ratios is not due to field stratification (i.e., majority-women fields have both more women in the audience and more women speakers; empirical $p < 0.004$ by permutation of questions across speakers within a subfield). The effect of speaker gender on question askers is also detectable in data recorded live at Biology of Genomes 2015–2018 and in data recorded live at ASHG 2015–2016 (Figures S2A and S2B).

Measurable Effect of Public Discussion as an Intervention to Increase Women's Participation

Developing a baseline measure of representation provides a basis for comparison after interventions to increase the participation of women in question-and-answer sessions.

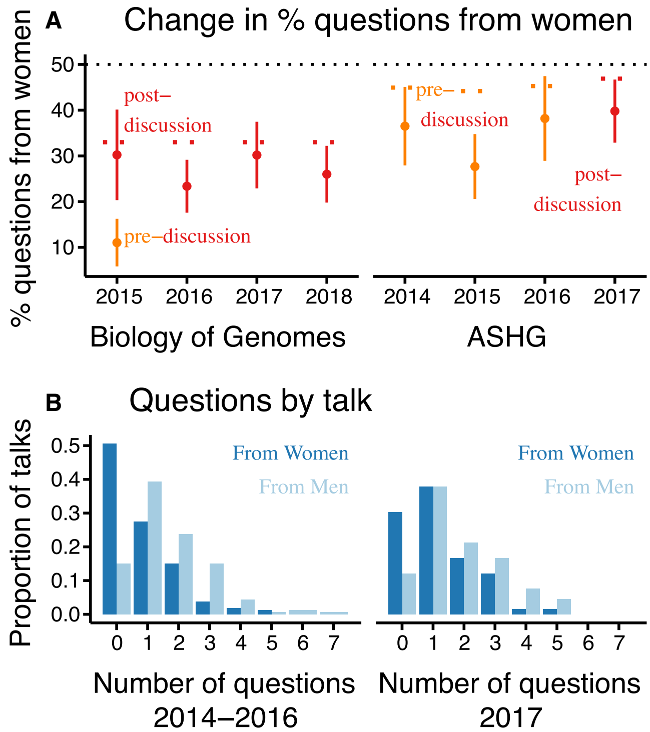

Figure 4. Public Discussion Increases Participation of Women in Question-and-Answer Sessions
(A) Percent of questions from women over time at two annual meetings in genetics (Biology of Genomes, "BoG" and the American Society of Human Genetics, "ASHG"). 804 questions at BoG were recorded live at the meeting; 666 questions at ASHG were recorded via webcast. Each point represents the proportion of questions asked by women in a given year. Vertical lines represent ± 2 binomial standard errors. Points in orange show questions asked before meeting-wide discussions of the participation of women during question-and-answer sessions, points in red occurred after this intervention. Colored dotted segments represent the proportion of women speakers at the corresponding meeting. Black dotted line marks equal participation (50%).
(B) Histogram of the number of questions per talk at ASHG. Questions asked by women are in dark blue, men in light blue. Left panel shows data from 160 talks from the 2014–2016 meetings (before public discussion of women's participation in question-and-answer sessions). Right panel shows data from 66 talks from the 2017 annual meeting (after intervention). All questions recorded via ASHG webcast.

In particular, we wondered whether alerting meeting attendees to the relative under-participation of women would impact the proportion of questions asked by women. We first noticed this effect at the Biology of Genomes (BoG) annual meeting in 2015. Midway through the meeting, N.T. shared several findings related to this work on Twitter. This sparked a meeting-wide discussion of the under-participation of women in question-and-answer sessions. This discussion culminated in an immediate policy change: the first question after every talk was thereafter offered to a trainee.

This discussion and the subsequent policy change was followed by a dramatic increase in the proportion of questions asked by women (Figure 4A). Interestingly, this effect persisted over time; in subsequent BoG meetings, the proportion of questions asked by women remained higher than it was prior to public discussion in 2015. Anecdotal evidence suggests that BoG attendees were still discussing this work and its implications among themselves, on Twitter, and with N.T. and E.C.G. in 2018. This may explain the persistent increase in the participation of women. However, this effect may also be driven, at least in part, by a broader evolution of perceptions of women in the workplace (e.g., the "MeToo" movement).

To test the broader applicability of this awareness-based intervention, we repeated this experiment at ASHG in 2017. In this case, N.T. described preliminary findings from this work in a plenary talk on the opening night of the annual meeting. The total proportion of questions asked by women following invited and plenary talks after the awareness-increasing intervention was similar to that seen in prior years (data collected via webcast; Figure 4A). However, the proportion of talks with zero questions from women decreased substantially (from 51% to 30%; Figure 4B). This suggests that there is a persistent effect of this intervention, across conferences.

The observed shift in the questions asked at both BoG and ASHG after awareness-increasing interventions suggests that public discussion of issues of diversity and participation is sufficient to alter behavior. There are several differences between the meetings, which we speculate may have impacted the performance of the intervention. BoG is a more intimate, single-track meeting, while ASHG is a large conference with many talks running simultaneously. Further, the intervention at ASHG was limited to discussion and was not accompanied by any policy change regarding question-and-answer sessions.

Crowd-Sourced Recording of Question-and-Answer Sessions

To help spread awareness of these findings at the 2017 ASHG annual meeting, after N.T.'s plenary talk, we released a web application through which meeting attendees could collect data regarding participation in question-and-answer sessions.

81 unique raters submitted 1,008 question records at 175 talks over the course of the meeting (participating raters submitted a median of 4 question records and a maximum of 109). 39% of all talks at the meeting had at least one question recorded and 24% had questions recorded by multiple raters (for a summary of the submitted data, see Figure S5).

We then aggregated question records submitted by multiple raters into a single, consensus set of questions for each talk (i.e., at a given talk, two or more raters may record the same question; these duplicate records were combined to avoid double-counting). To do this, we developed and applied a centroid-based approach that leverages the time at which a question record was submitted as well as an ID to anonymously track the submitting rater (a schematic is shown in Figure 5A). We use this approach to combine records from multiple raters representing single questions,

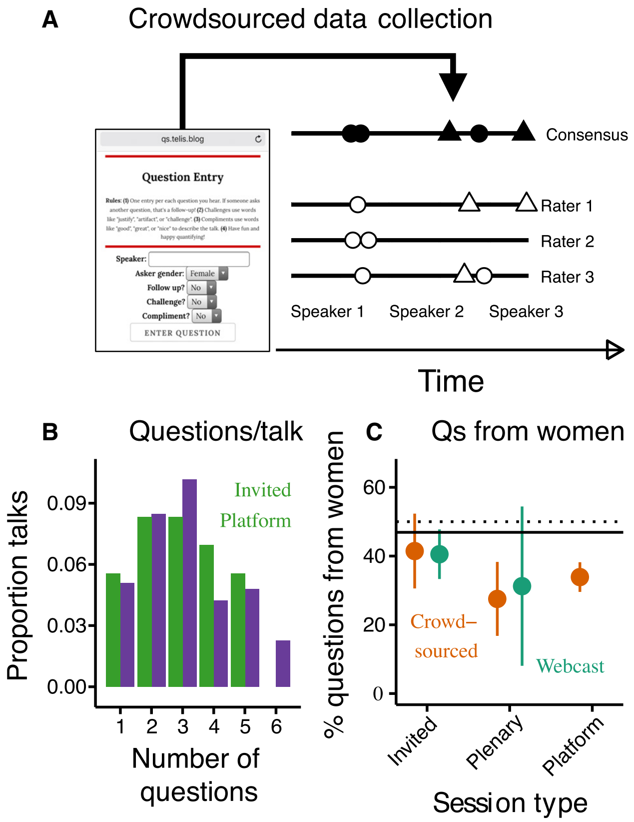

Figure 5. Crowd-Sourced Question Recording from ASHG 2017
Data are 1,008 question records submitted by 81 participating raters (collapsed into 612 consensus questions) at the 2017 ASHG annual meeting.
(A) Schematic of crowd-sourced data collection and processing. Publicly accessible web application portal for data entry shown on the left, example database entries (records) for a single talk on the right, consensus list of questions for that talk on the top right. Questions asked by men are shown as circles, those asked by women as triangles. Consensus questions are determined by collapsing question records from all raters following a single talk (see Material and Methods).
(B) Proportion of talks with a given number of consensus questions. Invited talks shown in green, platform talks in purple.
(C) Percent of questions asked by women in each session type at ASHG 2017. Estimates from 612 questions asked following 175 talks collected at the meeting via crowd-sourcing are in orange. Estimates from 201 questions asked following 66 talks recorded via the ASHG webcast are shown in green. Vertical lines represent ± 2 binomial standard errors. Dotted line marks equal participation (50%), solid line marks overall representation of women presenters at ASHG 2017 (47%).

focusing on limiting double-counting to avoid systematic bias in our estimation of participation. Rather than limiting our analysis to one record of the talk, the consensus approach allows us to maximize the amount of data we use from our crowd-sourced rater participants. We can confidently utilize the time of record submission, as the vast majority of records were submitted live as questions were asked at the meeting (Figure S4).

Using this approach, we collapsed the 1,008 question records into 612 consensus questions. For each talk, we detected fewer consensus questions than the average number of question records submitted per rater (see Figure S5), suggesting that many raters record only a subset of questions in a given question-and-answer session. As ASHG is a large conference with many sessions in parallel, many attendees move sessions between talks. This pattern of missing data is, therefore, unsurprising.

The proportions of questions asked by women at invited and plenary talks as inferred from our crowd-sourced consensus questions are very similar to those collected from the ASHG 2017 webcast (Figure 5C). This suggests that our crowd-sourcing app and post-processing provides a viable and scalable method for collecting observational data on question-asking behavior at large scientific conferences.

We were particularly excited by the 453 consensus questions (from 633 question records) recorded at 137 platform talks from 12 distinct topics. These data span a much broader portion of human genetics than we could have observed without a crowd-sourced effort. In addition, these data allow us to compare the gender ratio of questions following platform talks (programmed from submitted abstracts), where audiences tend to be smaller and speakers more junior, to that after invited talks, where audiences are larger and speakers tend to be more senior. Our observation that the overall proportion of questions asked by women was similar between platform and invited talks (Figure 5C) suggests that the seniority of the presenter is not a defining factor in the gender ratio of questioners.

We believe that the presentation of our data in a plenary talk, combined with crowd-sourced data collection at ASHG, increased attendee engagement with this topic and helped spread the awareness that led to increased participation of women (in terms of the number of talks with at least one question from a woman). Further, the data collected via crowdsourcing provided valuable additional insight into differences in participation between men and women across subfields and session types.

Discussion

This work highlights several hurdles on the path to achieving gender equity in science. First, coarse-grained diversity metrics may conceal persistent gender disparities. We see dramatic variation in the representation of women across subfields at the American Society of Human Genetics (ASHG), despite equal overall representation. Even as we approach broad gender parity, some women still find themselves in heavily male-dominated subfields.

Second, disparities in the participation of men and women often exceed disparities in representation. At ASHG, this is true for oral presentations and for question-and-answer sessions. Importantly, this trend holds in majority-women subfields. Together, these findings suggest that achieving equal representation at a meeting, or in a field, may fall short of achieving true equity.

Third, gendered differences in participation are at least partially attributable to factors internal to women scientists rather than to external biases. Question asking is voluntary, self-initiated, and, at least at ASHG, requires no external review. Yet, women choose to ask fewer questions than men. This suggests that top-down interventions, such as blind review, are unlikely to fully address the under-participation of women at academic meetings.

Finally, many at ASHG and elsewhere have asked the authors of this paper how we can increase the participation of women in question-and-answer sessions. ASHG and other scientific societies have played a key role in promoting scientific training and communication by their members;[29] the last decade has seen a large increase in trainee-oriented programs. We believe such efforts are excellent opportunities to engage with and improve diversity in our community. And, while we are unable to claim that increased question asking correlates with increased career success for women scientists, we do see that, following wide public discussion of the relative under-participation of women, women choose to ask more questions. This response suggests that women see value in increased question asking at conferences.

Further research is required to more fully understand the different perspectives and/or motivations of men and women with respect to question asking at academic conferences. It would also be valuable to evaluate the effect of question asking on career outcomes. More importantly, all our current work has focused on disparities across one measure (gender identity) measured imperfectly by external perception. Future work could and must explore a broader self-reported and non-binary representation of gender, as well as the effects of race, ethnicity, and other demographic characteristics on meeting participation.

While much work remains to increase diversity in the sciences, we conclude that crowdsourced data collection engages the community in issues of diversity, promotes visibility of such issues at academic meetings, and opens up data-driven explorations of variable representation and participation in our field. We encourage other researchers to make their academic communities partners in understanding and promoting diversity in STEM.

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2019.06.004.

Acknowledgments

The authors would like to thank the presenters and attendees at ASHG 2014–2017 and BoG 2015–2018, where we collected the data for this project. We would also like to thank the anonymous raters who participated in our crowd-sourced data collection at ASHG 2017, as well as the ASHG staff for sharing conference data for validation of our pipeline. In addition, we would like to thank Benjamin Lehmann for his instrumental contributions to the web application for crowd-sourced data collection. Finally, we would like to thank the members of the Pritchard lab and others in the Stanford community who have participated in many spirited discussions of this work. This research was supported by NIH R01 HG008140 and HG009431 (to J.K.P.) and NIMH 2P50 MH100029 (to C.G.). C.G. acknowledges support from the Fondation Brocher.

Declaration of Interests

E.C.G. and C.G. have served in advisory roles on ASHG committees, including the Program Committee (C.G.), the Communications Committee (E.C.G., C.G.), and the Social Issues Committee (C.G.). J.K.P. served as an organizer of BoG in 2016–2018.

Received: December 18, 2018
Accepted: June 5, 2019
Published: June 27, 2019

References

1. Holman, L., Stuart-Fox, D., and Hauser, C.E. (2018). The gender gap in science: How long until women are equally represented? PLoS Biol. *16*, e2004956.
2. van den Besselaar, P., and Sandström, U. (2017). Vicious circles of gender bias, lower positions, and lower performance: Gender differences in scholarly productivity and impact. PLoS ONE *12*, e0183301.
3. West, J.D., Jacquet, J., King, M.M., Correll, S.J., and Bergstrom, C.T. (2013). The role of gender in scholarly authorship. PLoS ONE *8*, e66212.
4. Murray, D., Siler, K., Lariviere, V., Chan, W.M., Collings, A.M., Raymond, J., and Sugimoto, C.R. (2018). Gender and international diversity improves equity in peer review. biorXiv.. https://doi.org/10.1101/400515.
5. Macaluso, B., Larivière, V., Sugimoto, T., and Sugimoto, C.R. (2016). Is science built on the shoulders of women? A study of gender differences in contributorship. Acad. Med. *91*, 1136–1142.
6. Lerback, J., and Hanson, B. (2017). Journals invite too few women to referee. Nature *541*, 455–457.
7. Larivière, V., Ni, C., Gingras, Y., Cronin, B., and Sugimoto, C.R. (2013). Bibliometrics: global gender disparities in science. Nature *504*, 211–213.
8. Shishkova, E., Kwiecien, N.W., Hebert, A.S., Westphall, M.S., Prenni, J.E., and Coon, J.J. (2017). Gender diversity in a STEM subfield - Analysis of a large scientific society and its annual conferences. J. Am. Soc. Mass Spectrom. *28*, 2523–2531.
9. Kalejta, R.F., and Palmenberg, A.C. (2017). Gender parity trends for invited speakers at four prominent virology conference series. J. Virol. *91*, 1–17.
10. Schroeder, J., Dugdale, H.L., Radersma, R., Hinsch, M., Buehler, D.M., Saul, J., Porter, L., Liker, A., De Cauwer, I., Johnson, P.J., et al. (2013). Fewer invited talks by women in evolutionary biology symposia. J. Evol. Biol. *26*, 2063–2069.
11. Débarre, F., Rode, N.O., and Ugelvig, L.V. (2018). Gender equity at scientific events. Evol Lett *2*, 148–158.
12. Isbell, L.A., Young, T.P., and Harcourt, A.H. (2012). Stag parties linger: continued gender bias in a female-rich scientific discipline. PLoS ONE *7*, e49682.

13. Johnson, C.S., Smith, P.K., and Wang, C. (2017). Sage on the stage: women's representation at an academic conference. Pers. Soc. Psychol. Bull. *43*, 493–507.
14. Jones, T.M., Fanson, K.V., Lanfear, R., Symonds, M.R.E., and Higgie, M. (2014). Gender differences in conference presentations: a consequence of self-selection? PeerJ *2* (e627), e627.
15. Epstein, N., and Fischer, M.R. (2017). Academic career intentions in the life sciences: Can research self-efficacy beliefs explain low numbers of aspiring physician and female scientists? PLoS ONE *12*, e0184543.
16. Käfer, J., Betancourt, A., Villain, A.S., Fernandez, M., Vignal, C., Marais, G.A.B., and Tenaillon, M.I. (2018). Progress and prospects in gender visibility at SMBE annual meetings. Genome Biol. Evol. *10*, 901–908.
17. Ford, H.L., Brick, C., Blaufuss, K., and Dekens, P.S. (2018). Gender inequity in speaking opportunities at the American Geophysical Union Fall Meeting. Nat. Commun. *9*, 1358.
18. Carter, A.J., Croft, A., Lukas, D., and Sandstrom, G.M. (2018). Women's visibility in academic seminars: Women ask fewer questions than men. PLoS ONE *13*, e0202743.
19. Hinsley, A., Sutherland, W.J., and Johnston, A. (2017). Men ask more questions than women at a scientific conference. PLoS ONE *12*, e0185534.
20. Davenport, J.R.A., Fouesneau, M., Grand, E., Hagen, A., Poppenhaeger, K., and Watkins, L.L. (2014). Studying gender in conference talks – data from the 223rd meeting of the American Astronomical Society. arXiv. arXiv, 1403.3091.
21. Schmidt, S.J., and Davenport, J.R.A. (2017). Who asks questions at astronomy meetings? New Astron. *1*, 1–2.
22. Pritchard, J., Masters, K., Allen, J., Contenta, F., Huckvale, L., Wilkins, S., and Zocchi, A. (2014). Asking gender questions: Results from a survey of gender and question asking among UK Astronomers at NAM2014. arXiv. https://doi.org/10.1093/astrogeo/atu245.
23. Campbell, J., and Mayer, R.E. (2009). Questioning as an instructional method: does it affect learning from lectures? Appl. Cogn. Psychol. *23*, 747–759.
24. Harper, K.A., Etkina, E., and Lin, Y. (2003). Encouraging and analyzing student questions in a large physics course: meaningful patterns for instructors. JRST *40*, 776–791.
25. Van Der Meij, H. (1994). Student questioning: A componential analysis. Learn. Individ. Differ. *6*, 137–161.
26. Brescoll, V.L. (2011). Who takes the floor and why. Adm. Sci. Q. *56*, 622–641.
27. Moss-Racusin, C.A., and Rudman, L.A. (2010). Disruptions in women's self-promotion: The backlash avoidance model. Psychol. Women Q. *34*, 186–202.
28. Wais, K. (2016). Gender prediction methods based on first names with genderizeR. R J. *8*, 17–37.
29. Braha, J. (2017). Science communication at scientific societies. Semin. Cell Dev. Biol. *70*, 85–89.

The American Journal of Human Genetics, Volume *105*

Supplemental Data

Public Discussion Affects Question Asking at Academic Conferences

Natalie Telis, Emily C. Glassberg, Jonathan K. Pritchard, and Chris Gunter

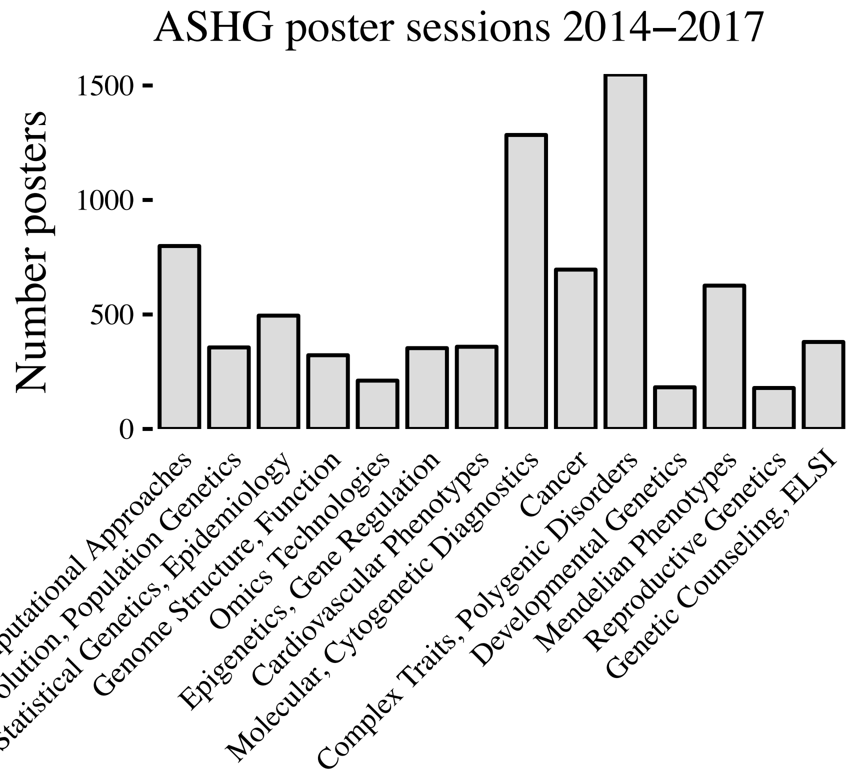

Figure S 1: **Number of poster presentations by subfield**. Data are 11,351 abstracts of poster presentations from the 2014-2017 annual meetings of the American Society of Human Genetics. Abstract information was mined from publicly available abstract books; for each abstract, the subfield is the topic category to which it was submitted. Subfields (x-axis) are ordered by the proportion of women poster presenters.

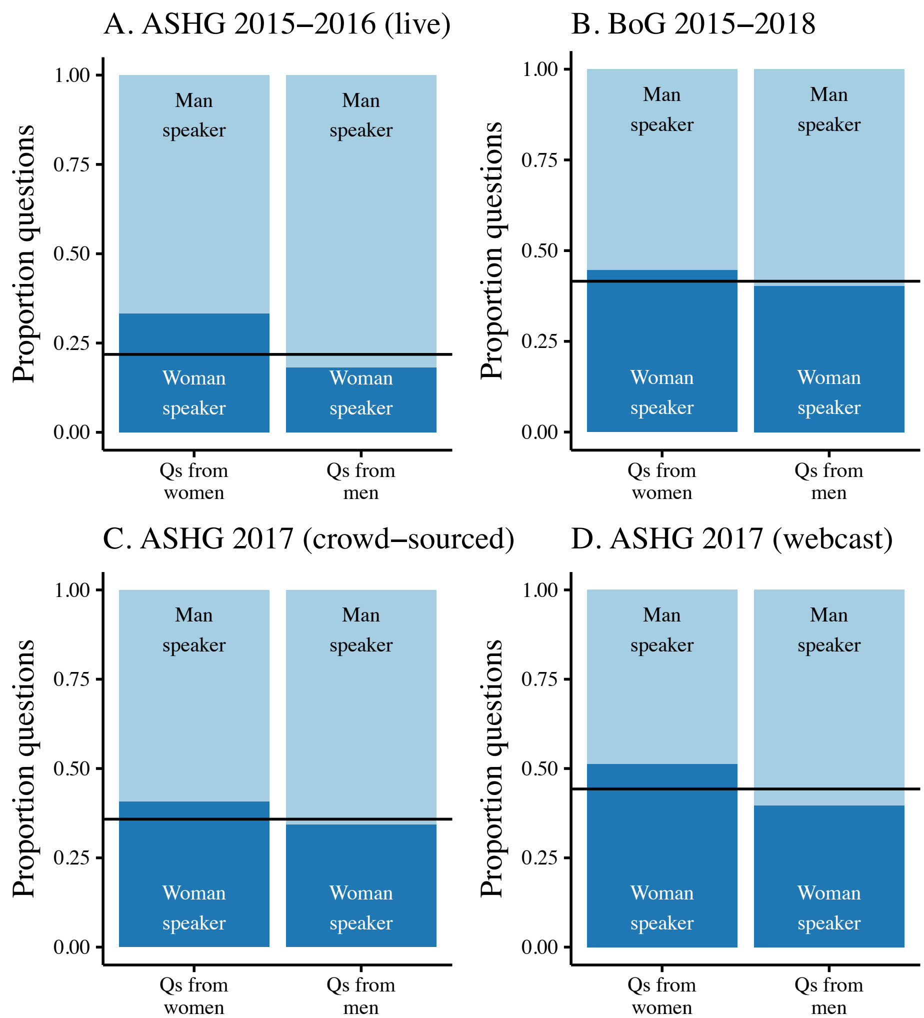

Figure S 2: **Replication of the effect of speaker gender on proportion of questions from women**. Filled bars show the proportion of questions from women and men, respectively (x-axis) addressed to women (dark blue) and men (light blue) speakers. Black line indicates the overall proportion of women speakers (and, therefore, the expected proportion of questions addressed to women). (A) Data are 168 questions following 58 talks at the 2014-2016 ASHG annual meeting (questions recorded live during meeting attendance). (B) Data are 635 questions following 166 talks at Biology of Genomes 2015-2018 (questions recorded live during meeting attendance). (C) Data are 201 questions following talks from invited and plenary talks at ASHG 2017 (questions recorded via webcast). (D) Data are 612 questions following 175 talks at the 2017 ASHG annual meeting (questions recorded by crowd-sourcing and, for each talk, records collapsed into a consensus set of questions asked).

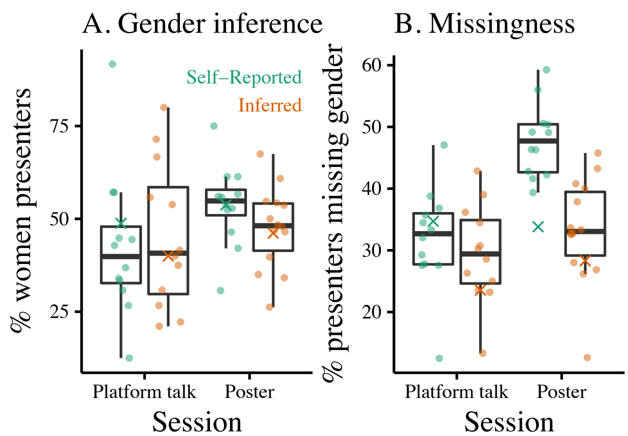

Figure S 3: **Automated inference of gender proportions recapitulates self-reported gender proportions**. (A) For each of 3,092 presentations from the 2017 ASHG annual meeting (and 3,472 presentations from the 2016 ASHG annual meeting), gender was inferred using first-author first name and the package `genderizeR` (see Materials and Methods in main text). Each orange point shows, in 2017, the proportion of presenters in a subfield (defined by a session topic to which abstracts were submitted) that were computationally inferred to be women. Each green point shows, in 2017, the proportion of presenters who self-reported as women in a subfield (data provided by ASHG staff). Boxplots show quantiles of the proportion of women presenters across subfields. Xs show the proportion of all presenters who were inferred to be, or self-reported as women in 2016. (B) The proportion of presenters for whom gender cannot be inferred (orange), or who choose not to report their gender (green) for each subtopic in 2017 (points) and overall in 2016 (Xs).

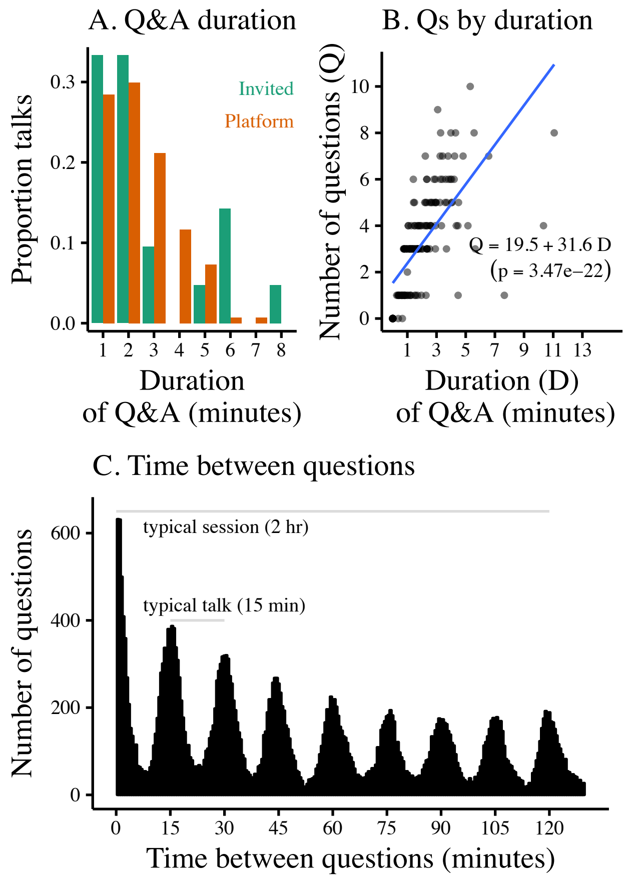

Figure S 4: **Timing of crowd-sourced question recording**. Data are 632 questions inferred from crowd-sourced data collection following 175 talks at the 2017 ASHG annual meeting. The duration of a Q&A was determined to be the difference between the first and last question-record submissions. (A) Histogram of the duration of question-and-answer sessions. Durations of Q&As following 30-minute invited talks are shown in green, those following 15-minute platform talks are in orange. This figure excludes six talks with question-and-answers that lasted longer than 8 minutes (perhaps due to question-records being submitted late, rather than during the session). (B) Relationship between the duration of a question-and-answer and number of questions. Each point represents a single talk. The number of questions is determined by combining information from all raters for a single talk into a single, consensus set of questions asked (see Materials and Methods in main text). Best-fit line of the relationship between the number of questions and the duration of the Q&A is shown in blue. This analysis excludes three talks with question-and-answer sessions that lasted longer than 13 minutes. (C) Histogram of the distance between all pairs of question-records. The periodicity reflects the structure of oral presentation sessions; many records are submitted simultaneously following a talk and many records are submitted 15 minutes later, at the conclusion of the subsequent 15 minute talk, etc. This pattern persists through the duration of a typical, 2hr long session. This suggests that most question-records are submitted as questions are being asked, allowing us to use the timestamp of record submission when determining consensus questions.

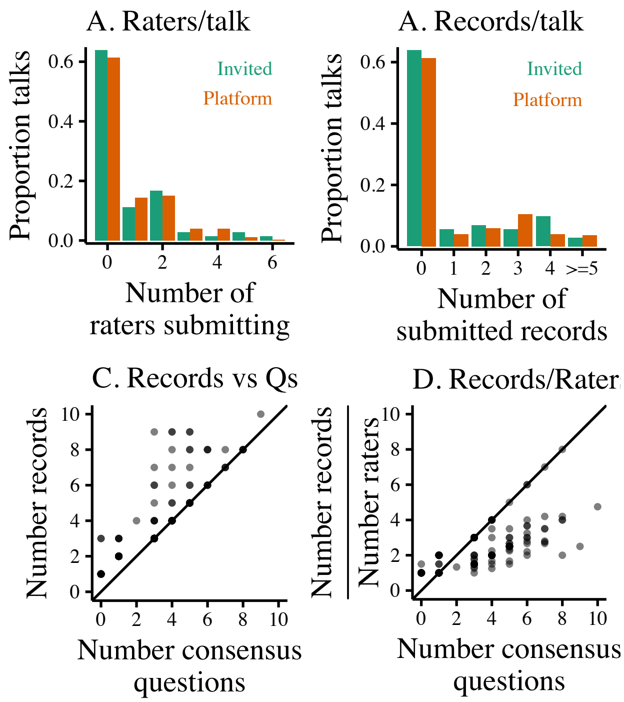

Figure S 5: **Processing of crowd-sourced question data from ASHG 2017**. Data are 1,008 question-records submitted by 81 participating raters (collapsed into 612 consensus questions) at the 2017 ASHG annual meeting. (A) Proportion of talks with questions recorded by a given number of participating raters. (B) Proportion of talks with a given number of unique, submitted question-records. Note that talks with zero records include missing data (talks with no raters present) and talks at which no questions were asked. Invited talks shown in green, platform talks in orange. (C) For each talk, the number of submitted question-records (y-axis) compared to the number of consensus questions (x-axis). Consensus questions were determined by collapsing all submitted records for each (see Materials and Methods in main text). Black line marks $y = x$. (D) For each talk, the number of consensus questions (x-axis) compared to the number of submitted question-records **per unique rater** (y-axis). Black line marks $y = x$. There are generally more records than consensus questions, but fewer than fewer records per rater than consensus questions, suggesting both that multiple raters often record the same question and that raters often fail to record every question at a talk.

		2014	2015	2016	2017	2018
ASHG	**Posters**	2964	2670	3069	2648	—
	Talks	470	470 (31)	403 (27)	444	—
	Webcast	55	65	40	66	—
	Questions	153	183 (84)	129 (84)	201 (175)	—
BoG	**Talks**	—	50	56	41	58
	Questions	—	231	214	159	201 (612)

Table S 1: **Sample sizes (presentations and questions) over four years of conference attendance**. 'ASHG' represents data collected at annual meetings of the American Society of Human Genetics. For ASHG, 'Talks' represents the total number of talks that occured at the meeting, while 'Webcast' represents the talks available via online webcast (invited and plenary talks). 'Questions' represents data recorded via webcast. Talk and question numbers in parentheses represent data recorded live during meeting attendance. For 2015-2017, these data were collected by ECG and NT, for 2017, these data were collected via crowd-sourcing. 'BoG' represents data collected at Biology of Genomes. For BoG, all talks were attended and questions recorded live during meeting attendance.

Caption for Table S 2: **Recorded talk sessions and topics**. Years, session names, and talk types (Platform and Plenary sessions by application, as well as Invited Scientific, Invited Education, and Invited Social Issues) were collected from the abstract books from the 2014-2017 annual meetings of the American Society of Human Genetics. For platform sessions, 'Topic' shows the topic categories to which abstracts were submitted. Note that some sessions were composed of talks submitted to multiple topic categories (labelled 'Mixed'). For invited sessions, topics represent groupings of poster session topics, and are manually assigned based on the titles of the talks included in each session. 'Data' detail how question data were collected (via recorded webcast, by ECG/NT live at the meeting, or crowd-sourced by participating raters via a publicly available web application).