

Gender-Related Systematics in the NRAO
Proposal Review Process

Update Including all Proposals from Cycles 12A-19A

Gareth Hunt, Frederic R. Schwab and Lewis Ball
NRAO

25 February 2019

1 Introduction

In 2016, Lonsdale et al. [LSH] undertook a study of gender systematics in the NRAO and ALMA proposal review processes. Reid [RHST] had previously performed a study on the same subject for the Hubble Space Telescope (HST) proposal system. A significant gender-related effect (in favor of proposals with male PIs over those with female PIs) was found in the ALMA process, and a similar effect was found for the NRAO instruments, but to a lesser extent and with some reversals in the trend of male advantage, when examined by telescope and over time.

We have continued to monitor the gender systematics since the publication of that study. This note is a simplified update to include subsequent proposal cycles of the AUI/NRAO telescopes - the Green Bank Telescope (GBT), the Very Large Array (VLA) and the Very Long Baseline Array (VLBA). We do not here address subsequent ALMA reviews. We also do not address other potentially significant parameters such as PI seniority/prestige, geographic origin, and review panel science field.

The original paper [LSH] included results for the NRAO review process from 2012 (12A) through 2016 (16A). This note extends the review samples through 2019 (19A).

| Year | GBT | VLA | VLBA | All | %F SRP |
|-------|-----|-----|------|-----|--------|
| 2012 | | | | | 25.0 |
| 2013 | F | | | f | 26.0 |
| 2014 | M | | | M | 27.1 |
| 2015 | | | | | 21.9 |
| 2016 | | m | | m | 21.0 |
| 2017 | | | | M | 26.2 |
| 2018 | m | f | | | 44.0 |
| 2019A | | | | | 49.1 |
| All | m | M | | M | 28.3 |

Table 1: This is a simplified summary by year showing just the extreme cases of gender imbalance using an imbalance key (see section 2 for details). Both semesters in a year have combined for the individual telescopes as was done in [LSH] for comparison with other telescopes, such as ALMA and the HST.

2 Simplified presentation, Anderson-Darling

The two distributions of scores awarded to female-PI and male-PI proposals were compared using several statistics. In this section, we present a summary of the Anderson-Darling p-values (ADp). A low ADp indicates that the two distributions are not from a common parent distribution; a higher p-value is indicative of a degree of commonality. In order to compare the female-PI and male-PI distributions in the tables, we assign an imbalance key (IK) indicating which gender, if any, is favored:

- an **upper case letter (F or M)** if $ADp \leq 0.1$ (i.e., **strongly distinct**, i.e., $>\sim 90\%$ confidence). When there is a significant imbalance, the ADp does not indicate which distribution has the higher scores. This is determined readily by reviewing the data plots (section 3) to give **F** for imbalance in favor of proposals with female-PIs, **M** for male-PIs.
- a **lower case letter (f or m)** if $0.1 < ADp \leq 0.2$ (**distinct**, i.e., $>\sim 80\%$ confidence). The letter is determined as above.
- blank otherwise

Our assignment of confidence levels based on the ADp values is not rigorously justified, but is consistent with [LSH]; Babu and Feigelson [BF] recommend confirmation by bootstrap resampling, which has not been

done for Anderson-Darling analysis. However, the confidence intervals for our quartile score analysis of section 3 are computed directly via bootstrapping.

Tables 1 and 2 present a simplified summary equivalent to table 3 in the original paper [LSH].

It should be noted that most individual ADp values do not show a statistically significant imbalance by semester or when combined to be presented by year. However, as mentioned in [LSH], the combination of all results shows a significant imbalance in favor of proposals with male PIs.

Note, for example in 2017, there was no significant imbalance for any individual telescope, but when considered together there was a significant indication of male advantage. Note also that the VLBA does not show any gender imbalance for individual semesters; the results are not statistically significant as the data are too sparse. For the other telescopes and for the combination of all three, there is a significant gender-related advantage for proposals with male PIs.

| Semester | IK | ADp | %F PI | %F SRP |
|----------|----|------|-------|--------|
| 12A | | 0.86 | 26.9 | 25.0 |
| 12B | | 0.40 | 26.2 | 25.0 |
| 13A | | 0.34 | 26.0 | 25.0 |
| 13B | | 0.51 | 24.8 | 27.1 |
| 14A | | 0.57 | 30.2 | 29.2 |
| 14B | M | 0.05 | 30.3 | 25.0 |
| 15A | | 0.73 | 24.8 | 18.8 |
| 15B | | 0.49 | 27.4 | 25.0 |
| 16A | | 0.65 | 30.2 | 25.0 |
| 16B | M | 0.02 | 22.6 | 17.0 |
| 17A | m | 0.12 | 32.7 | 20.8 |
| 17B | | 0.25 | 30.6 | 31.5 |
| 18A | | 0.22 | 32.1 | 38.9 |
| 18B | F | 0.06 | 33.6 | 49.1 |
| 19A | | 0.73 | 32.0 | 49.1 |

Table 2: A summary by semester of the Anderson-Darling p-values comparing the distributions of the scores of proposals with female PIs and those with male PIs for all telescopes combined. Also included are the imbalance key (see section 2), the percentage of proposals with female PIs, and the percentage of female reviewers.

2.1 Reviewer gender ratio

In recent years, the NRAO has implemented a policy requiring consistent effort on behalf of the recruiters to populate the science review panels (SRP) with a gender ratio roughly equal to the astronomical community that we serve. This changing ratio is recorded in tables 1 and 2. Furthermore, the recruiters attempt to have at least two female reviewers on each SRP. Both goals have been consistently achieved since 2018.

As noted in [LSH], there is no clear trend comparing the gender-based success of proposals to the percentage of female reviewers, nor to the continuation pattern of SRP members.

3 Graphical presentation, quartile score analysis

We modeled distributions of the quartiles of the normalized scores by bootstrap replication and resampling (see [E] section 5.3). A series of graphs and figures were generated for each year combining proposal cycles A and B for the individual telescopes and for the combination of all three. These graphs and figures are also generated for each semester for the combination of all three. The full set of these plots is available on-line.¹ In this section, we have extracted just the modeled distributions and combined them into three figures.

¹<https://science.nrao.edu/science/reports/StatisticalData>

Figure 1 shows the quartile plots for each year, an extension of figure 2 in [LSH]. Note that most years show a higher score distribution for proposals with a male PI. The exception is 2013, consistent with the IKS in table 1.

Figure 2 shows the same information for semesters 2012A-2015B, and figure 3 for 2016A-2019A. Note that semesters 13A, 13B, 15A, 18B and 19A show signs of a higher score distribution for proposals with a female PI. However, of these only 18B showed a clear imbalance (see table 2). Semesters 14B, 16B and 17A, which show a clear imbalance in table 2 also show clear signs of a higher score distribution for male PIs.

Since the Anderson-Darling method matches the complete distribution, giving higher weight to the edges of the distribution, the results compared to other methods are not necessarily intuitive.

4 Conclusions

The results reported here, extending the analysis of [LSH] to semester 2019A, show that the outcomes of the NRAO proposal review process tend to favor proposals from male PIs over those from female PIs. There are insufficient data to identify trends.

AUI/NRAO is concerned by the gender-related imbalance revealed by these studies. AUI/NRAO is committed to a fair and equitable proposal review and time allocation process, and actively emphasizes to all reviewers each cycle that rankings and decisions must reflect only scientific merit, technical feasibility, and operational constraints. Since 2017, the NRAO has worked to raise awareness of these issues with reviewers and its user community, and has successfully achieved more balanced gender representation on its review panels. The Observatory is committed to delivering greater fairness in future, and is monitoring developments being implemented elsewhere. The NRAO may change its review processes in the future if there is compelling evidence to support doing so.

5 References

- LSH** Carol J. Lonsdale, Frederic R. Schwab and Gareth Hunt, *Gender-Related Systematics in the NRAO and ALMA Proposal Review Processes*, <http://arxiv.org/abs/1611.04795>, 16 November 2016.
- RHST** I. N. Reid, *Gender-based Systematics in HST Proposal Selection*, PASP, 126, 923, 2014.
- BF** G.J. Babu and E.D. Feigelson, *Goodness-of-fit and all that!*, Astronomical Data Analysis Software and Systems XV, ASP Conference Series #351, 127, 2006.
- E** Bradley Efron, *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics, 1982.

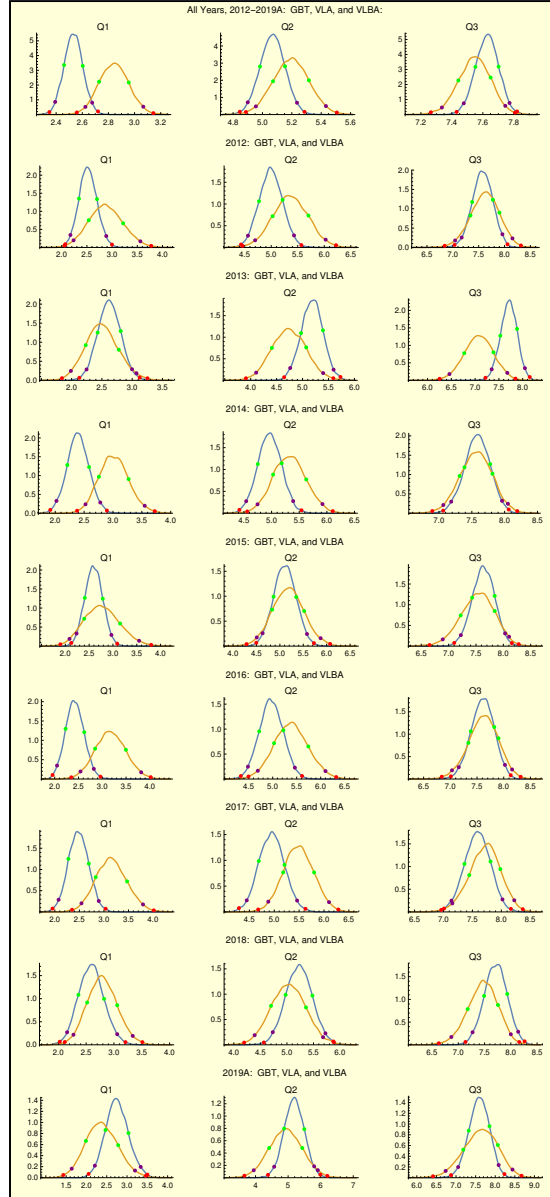


Figure 1: Modeled distributions of the 25th, 50th and 75th percentile of the normalized scores of proposals submitted to all telescopes derived by bootstrap resampling. Orange curves: proposals with female PIs; blue curves: male PIs. Green, purple and red dots delimit, respectively, the 68%, 95% and 99% probability intervals. Top: total; subsequent rows are for years 2012-2018,2019A.

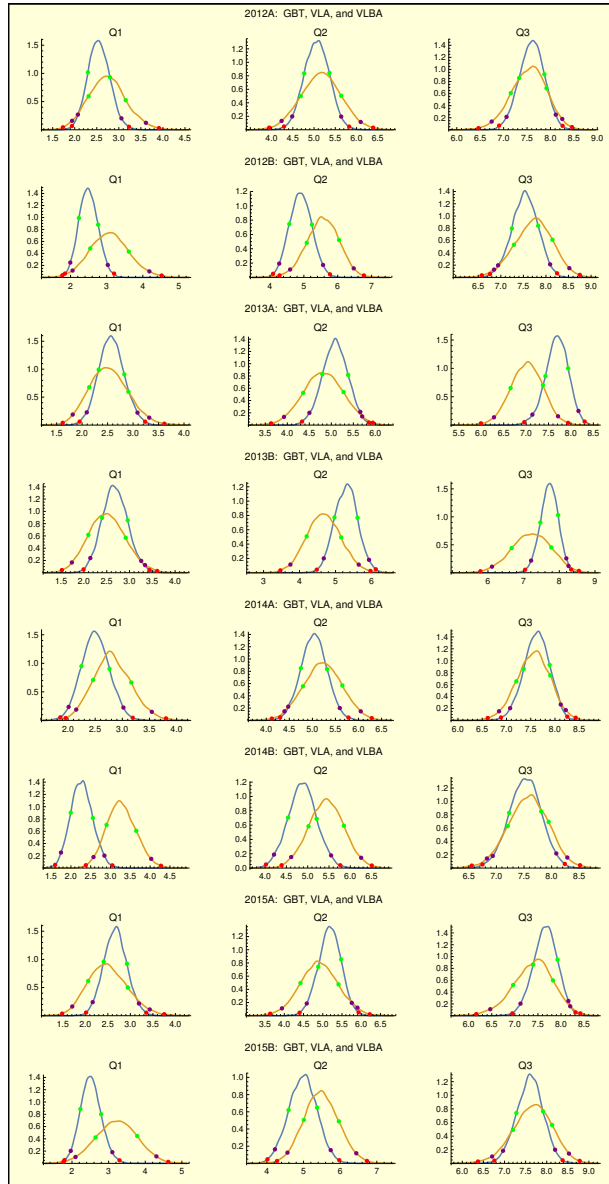


Figure 2: As for Figure 1. Semesters 2012A-2015B.

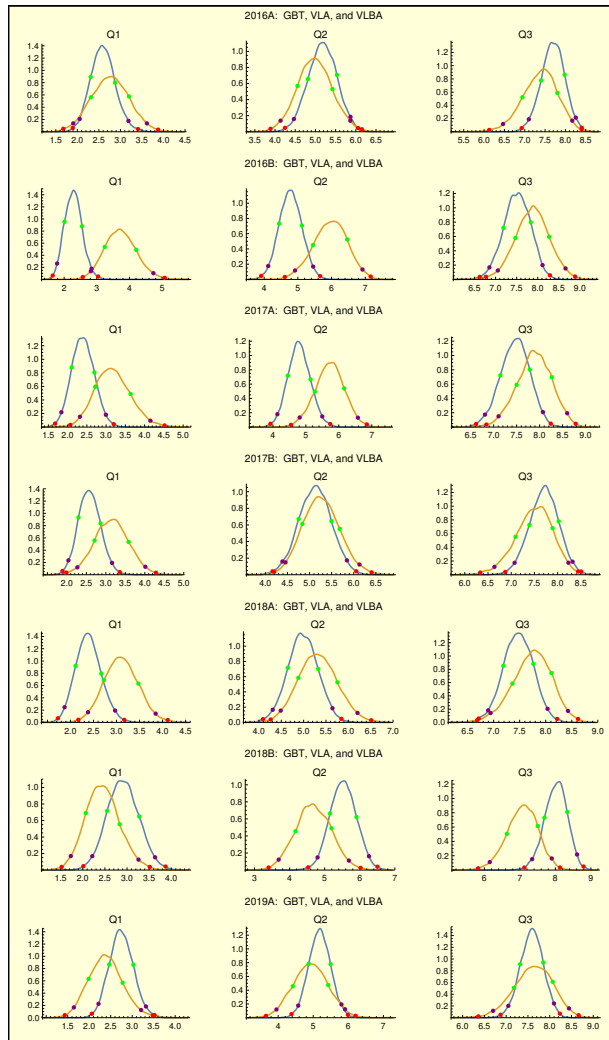


Figure 3: As for Figure 1. Semesters 2016A-2019A.