

## Astro2020 Activities, Projects, or State of the Profession Consideration White Paper

### Adopting Dual-Anonymous Practices in the Reviews for Resource Allocation in Astronomy

**Type of Activity:** ☐ Ground Based Project   ☐ Space Based Project  
☐ Infrastructure Activity   ☐ Technological Development Activity  
☒ State of the Profession Consideration   ☒ Other: *Recommendations to agencies/organizations*

**Principal Author:**

Name: Louis-Gregory Strolger

Institution: Space Telescope Science Institute

Email: strolger@stsci.edu

Phone: (410) 388-1348

**Co-authors:** I. Neill Reid (STScI)

**Abstract:** In an effort to reduce biases, Space Telescope Science Institute (STScI) adopted a dual-anonymous review in its time allocation process in which the identities of both the proposers and the reviewers are withheld until after the science program is selected. The results of the change (after two cycles of implementation) have so far been encouraging, with the success rates of female PIs no longer consistently lagging below the success rates of their male counterparts. Coupled with other examples like the large increase in the fraction awards to new investigators in Cycle 27, and the increasing trend in the fraction of submissions led by women, the adoption of the dual-anonymous review is poised to increase the participation of women and other underrepresented groups in the use of the *Hubble Space Telescope*, improving access to this highly impactful research resource.

STScI reached this decision after much consideration and consultation. We implemented changes in our peer-review time allocation process that borrow heavily from methods used in the anonymous review of manuscripts for journals. In this white paper we discuss some of the motivations for adopting a dual-anonymous process, overview the changes made to the time-allocation reviews, and present preliminary results and lessons learned. We also provide recommendations to other agencies and organizations to consider adopting similar processes in the peer-reviews for telescope time and other resource allocation.

## 1. THE IMPETUS FOR A DUAL-ANONYMOUS PROCESS

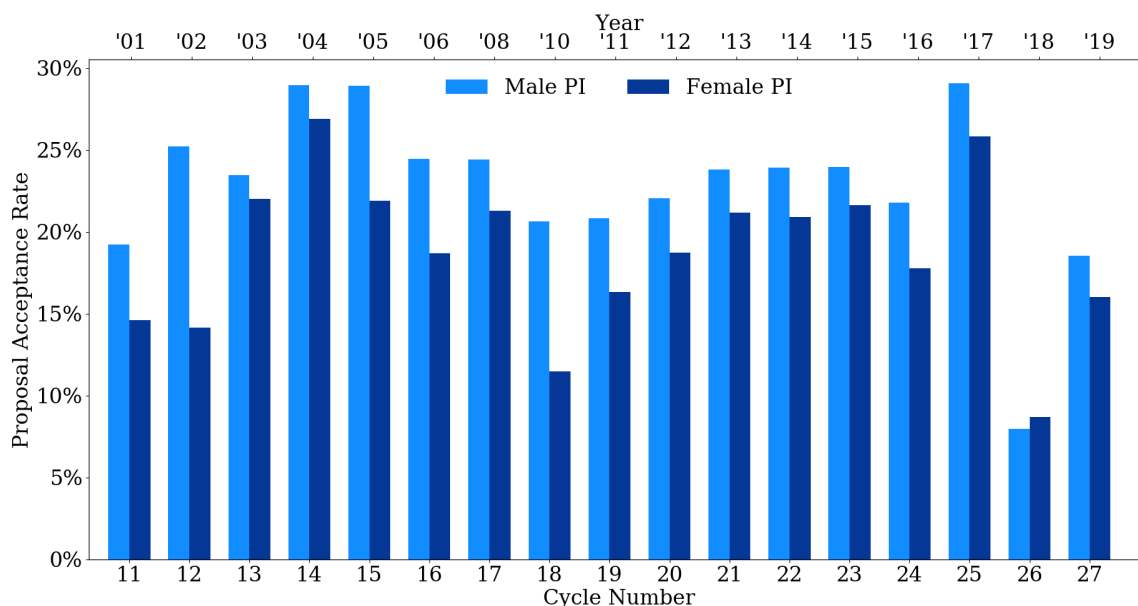
The STScI Directorate and Science Policies Group have actively sought ways of lessening the opportunity for conscious and unconscious bias in the *Hubble Space Telescope* (HST) time allocation (TAC) process. Yet, the results of HST reviews from 2001 to 2012 showed that the success rates of proposals lead by female principal investigators (19% on average) have been worse than those led by their male counterparts (23%). The imbalance is statistically significant with data from just a few observing cycles, and as Figure 1 shows is systemic, with female PIs falling short of male PIs year after year, seemingly indicating a continued presence of unconscious bias (Reid 2014).

Dual-anonymous<sup>1</sup> reviews are increasing in popularity within the science community (Kmietowicz 2008). Compared to singly-anonymous reviews, where the identities of the reviewers are hidden from the authors, dual-anonymous reviews add more anonymity to the process by hiding the identities of the authors from the reviewers. The objective is to reduce bias in the ratings of reviewers by removing demographic information about the authors that could trigger bias, such as gender, race, career stage, institution, or nationality. Despite suggestions that anonymous processes are ineffective at hiding the identity of authors, they are still highly regarded by many people in the scientific community (Mulligan, Hall, & Raphael 2013), and studies that have tested the accuracy of author “guessing” have concluded that these methods are often effective in protecting anonymity (Hill & Provost 2003, Justice et al. 1998).

Studies of the impact of dual-anonymous reviews are limited but promising. There is evidence that interventions that implement dual-anonymous over singly-anonymous processes decrease bias related to gender (Breda & Hillion 2016; Budden et al. 2008), institution, author prestige (Lee, Om, & Koh, 2000; Ross et al. 2006), age (Lee et al. 2016), and country of origin (Ross et al. 2006). The evidence of the efficacy in the reduction of bias should be reason enough to further implement dual-anonymous reviews.

In response to Reid (2014), STScI hired an independent consultant, Dr. Stefanie Johnson (Univ. of Colorado), who has worked with businesses and science groups in dealing with conscious and unconscious bias, to observe and assess the HST Cycle 25 peer-review process. In the first round

<sup>1</sup> The term “double blind”, which is often also used to refer to this type of process, is inadvertently ableist and devaluing people with a disability, and should be strictly avoided.



**Figure 1.** In each observing cycle, from 11 to 27 (2001- 2019), HST proposals led by men (light blue bars) have had better success rates than those led by women (dark blue bars), with the exception of Cycle 26 when the dual-anonymous process was first implemented. The gap in acceptance fraction in Cycle 27 is as small as it has been only 3 other times in the last 17 cycles, or over the last 19 years. Data source: STScI

of independent grading done prior to the in-person review (which sets the agenda for the in-person discussion), Johnson, and her then graduate student Jessica Kirk, found no evidence of gender bias in which proposals made it to the discussion stage. It was only in the in-person discussions that bias became evident, potentially because much of the in-person discussion on a given proposal focused on the track record of the proposers, rather than on the science presented in the proposal. Johnson and Kirk recommended implementing some anonymous process at the discussion stage to refocus the dialogue.

## 2. CHANGES IN THE REVIEW PROCESS ADOPTED AT STScI

A working group was charged with developing an implementation plan for the anonymous proposal reviews, starting with the HST Cycle 26 Delta TAC. The working group reviewed available literature, solicited opinions from the community, and presented their general framework and recommendations to the STScI Directorate and Science Policies Group.

**Changes in the instructions to proposers:** The process that STScI adopted took many lessons from the anonymous review of manuscripts for science journals. To facilitate such a review, proposals must be crafted from the start with anonymity in mind. Proposals should not include statements that directly identify the proposing team or participants within, including identifying self-references, or claim ownership of past work. This generally encourages proposers to write in a more passive voice, e.g., using phrases such as “As Doe et al. have shown. . .” as opposed to “As we have shown in Doe et al. . .”, and therefore requires some additional preparatory effort from investigators. STScI provides guidelines in its Calls for Proposals to help investigators adequately

anonymize their submissions. Proposers that do not make sufficient efforts to anonymize their submissions run the risk of having their proposal rejected for non-compliance.

**Changes in the instructions to reviewers:** TAC panelists are reminded that the primary objective of these reviews is to identify and select the most compelling science, not to reward the best science teams. Their evaluations at every stage of the review must be based on the scientific merit of the proposals, not the experience, expertise, or relative merit of the proposers. As such, the panelists are highly discouraged from delving into the identities of the investigators. STScI provides guidelines to reviewers to help focus their ratings and discussions.

**The role of levelers:** The practice of preserving anonymity in this type of review was expected to be unfamiliar to many in the community, as it is currently contrary to what is typically expected in similar reviews for other observatories and agencies. With this in mind, observers were stationed in the rooms of each panel with the sole objective of keeping the panel focused on the scientific merit of each proposal. These *levelers* are not listening for issues pertaining to the science, rather are focused on the discussion itself. If the discussion veers to comments on the proposing team, their past work, their validity, or their identities, the leveler's job is to refocus that discussion.

**Conflicts of interest:** A collateral benefit to dual anonymous reviews is that it greatly simplifies conflicts of interest issues— in principle, if the reviewer does not know who the proposers are then there could not be a conflict of interest. In practice, however, there are several interpersonal situations for which it is still prudent to track conflicts and avoid the perception of bias, even if identities are not evident. In our system, reviewers are deemed conflicted on proposals by family members and collaborators (or competitors) whom they have worked with (or against) in the last 3 years. STScI continues to track several of these relationships through its proposal database, the Astrophysics Data System (ADS), and through questionnaires upon volunteering for service on our TAC and other institutional reviews, and frequently declares conflicts for reviewers in advance. We require our panelists to declare any additional conflicts as they become aware of them. Reviewers are excluded from the discussion or ranking of proposals on which they are conflicted.

**A separate review of expertise:** As a public facility, HST time is openly available to any scientists who presents a highly compelling scientific case. However it is recognized that time is also a highly valued resource that must be used responsibly. The most frequently raised concern to the working group was that an anonymous process would make it difficult for the community to, in effect, police itself, maintaining that proposers should have demonstrable competency in the execution of science programs and the dissemination of those results.<sup>2</sup> In response to this concern,

<sup>2</sup> A corollary to this argument was that proposers for new time should be held accountable for past awards— past failures or delays in the timely publication of results would be perceived by reviewers as a weakness in requests for more time. Experience has shown, however, that this has not been the case.

the peer review now has an additional round of expertise review.

Proposers are be required to submit a Team Expertise and Background exposition with their proposal submissions. This free-form exposition is not anonymous, and is meant to cover work to be done, relevant expertise of team members, or a summary of results from relevant past programs (not just with HST). It could also contain a description of an analysis plan, including a description of the effort and the proposed roles of the investigators.

After the anonymous scientific ranking is complete, the review panel is given the list of investigators (alphabetized, with no indication of principal investigator) and the Team Expertise and Background sections for those proposals recommended for approval to the Director. If there are clear, compelling deficiencies in the expertise required to see through the goals of the proposal, panel must decide by consensus to flag the submission for potential disqualification, and provide a detailed justification in their comments to the Director.

### 3. EARLY RESULTS AND LESSONS LEARNED

It will likely take time to know if dual-anonymous process had indeed made an impact on the gender parity in the PI success rates. If a cycle is treated as a data point, inferential statistics such as the t-tests require at least a few cycles, if not several, to determine if there is a statistically significant difference in the gender groups. One of the specific recommendations of the working group was that it would, therefore, be beneficial to track other measures of success that, while possibly anecdotal, might better express the immediate impact of any changes.

In terms of these early successes, in the first cycle of implementation of this anonymous process (Cycle 26 Delta TAC) the proposal success rates for proposals led by women were, for the first time, higher than those led by men (at 8.7% and 8.0%, respectively). And while the trended imbalance returned in Cycle 27, the gap in acceptance fraction was as small as it has been only 3 other times in the 17 cycles held over the last 19 years (see Figure 1). The results of the Cycle 27 review also show that an incredible 30% of approved programs went to first-time principal investigators, compared with only 6% in Cycle 26, 21% in Cycle 25, and 5% in Cycle 24. Additionally, there has steady growth in the fraction of proposals led by women at a rate of about 0.5% per year (at around 28% in Cycle 27, up from only 19% in Cycle 11).

Levelers have been a highly beneficial addition to this review process. Many in the community are supportive of the dual-anonymous review, but as noted previously, the novelty of this application to has been a source of confusion for many reviewers. Levelers present in every review room allow for immediate responses to inquiries or actions which could easily break the anonymity in the review of a given proposal, leading to unfair results. Furthermore, as this is a highly competitive process, levelers reduce the tendency by panelists to manipulate the process toward personal goals.

The other major change was the addition of a specific review of expertise. Despite the large concern expressed to the working group, there's been considerable push back from our reviewers as to why we continue with this as part of the review. Many have stated that questions on competency are largely (if not completely) addressed in the quality of the scientific and technical justification in the proposals themselves, and they see little to the contrary in the expertise sections.

#### 4. RECOMMENDATIONS

**REC-1 Adopt dual-anonymous reviews for observing time.** *The adoption of a dual-anonymous process for HST has shown some successes thus far, and looks poised to improve the gender inequity in awarded time. Based on this success, the NASA Science Mission Directorate for Astrophysics Missions will adopt dual-anonymous practices in all of the General Observer programs within its missions, beginning in 2020. There is some evidence that observatories, especially those that are public facilities, are considering adopting similar practices. We highly encourage that these practices be widely adopted, especially at public facilities.*

**REC-2 Integrate a dual-anonymous process in other resource allocations.** *It is highly conceivable that these practices could be extended to other peer-review resource allocation processes, including those for money and jobs. In such practices, it may be prudent to implement a two-stage process, where the scientific merit is reviewed following an anonymous process, and the allocation of funds (or positions) is handled by a separate non-anonymous review. This may better utilizes the skills of scientists in reviewing science, and potentially effort, while leaving financial reviews to those with more appropriate experience.*

#### References

Breda, Thomas, and Mélina Hillion. "Teaching accreditation exams reveal grading biases favor women in male-dominated disciplines in France." *Science*, 353.6298 (2016): 474-478.

Budden, Amber E., et al. "Double-blind review favours increased representation of female authors." *Trends in ecology & evolution*, 23.1 (2008): 4-6.

Hill, Shawndra, and Foster Provost. "The myth of the double-blind review?: author identification using only citations." *Acm Sigkdd Explorations Newsletter*, 5.2 (2003): 179-184.

Justice, Amy C., et al. "Does masking author identity improve peer review quality?: A randomized controlled trial." *Jama*, 280.3 (1998): 240-242.

Kmietowicz, Zosia. "Double blind peer reviews are fairer and more objective, say academics." *BMJ: British Medical Journal*, 336.7638 (2008): 241.

Lee, Mushin, Kiyong Om, and Joon Koh. "The bias of sighted reviewers in research proposal evaluation: a comparative analysis of blind and open review in Korea." *Scientometrics*, 48.1

(2000): 99-116.

Mulligan, Adrian, Louise Hall, and Ellen Raphael. "Peer review in a changing world: An international study measuring the attitudes of researchers." *Journal of the Association for Information Science and Technology*, 64.1 (2013): 132-161.

Reid, I. Neill, "Gender-Related Systematics in HST Proposal Selection." *Publications of the Astronomical Society of the Pacific*, 126, 923

Ross, Joseph S., et al. "Effect of blinded peer review on abstract acceptance." *Jama*, 295.14 (2006): 1675-1680.