

# Report of the Working Group on Anonymizing Proposal Reviews

Lou Strolger (Chair, STScl), Peter Garnavich (Notre Dame), Stefanie Johnson (Leeds Business School, U. Colorado, Boulder), Mercedes Lopez- Morales (CfA, STUC), Andrea Prestwich (CfA), Christina Richey (JPL), Paule Sonnentrucker (STScl), Michael Strauss (Princeton), and Brian Williams (STScl). Tom Brown and Neill Reid (Ex-officios, STScl).

## Summary

The Working Group on Anonymizing Proposal Reviews was charged with developing a plan for implementing an anonymous proposal process at the Institute, beginning with the HST Cycle 26 TAC in 2018. In telecons over the period from January 24 to March 7, 2018, and an in-person meeting on March 28 & 29, 2018, we have considered the broad implications of a more fully anonymous proposal review system, reviewed literature and on-line examples of success in other scientific fields, and discussed input from colleagues here at the Institute and in the astronomy community. From these discussions we have developed a largely dual-anonymous peer review process, which includes a final check of team expertise and preparedness to further insure the feasibility of proposed science.

On an accompanying website, <https://outerspace.stsci.edu/display/APRWG>, we provide some responses to frequent and anticipated questions and comments, guidelines to proposers for how to develop their anonymized proposals, and guidelines for reviewers on what to do and not do in a dual-anonymous review. We expect that the Institute will continue to curate these documents, and integrate them into the proposal instructions and orientation materials to the TAC.

## Introduction

The Institute has actively sought ways of lessening the opportunity for conscious and unconscious bias in its proposal review process. Yet, the results of the reviews continue to show that proposals led by female PIs have a lower success rate than those led by male PIs (Reid 2014), which may indicate the continued presence of unconscious bias.

Double blind, or dual anonymous, 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 anonymity to the process by hiding the identities of the authors from the reviewers. The object is to reduce bias in

the ratings of reviewers by removing information about the authors that could give clues to demographic information such as gender, race, tenure, institution, or nationality. Despite suggestions that blinding interventions are ineffective at hiding the identity of authors, this process is highly regarded by many people in the scientific community (Mulligan, Hall, & Raphael, 2013). Furthermore, studies have tested the accuracy of author “guessing” and 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 blinding interventions that implement dual-anonymous over singly-anonymous 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). Evidence of the efficacy of blinding in the reduction of bias should be reason enough to implement dual-anonymous reviews at the Institute. Such processes would focus attention on the quality of the science and help create more equal opportunities for time with HST.

The Working Group on Anonymising Proposal Reviews (WGAPR, or WG hereafter) has been charged with developing an implementation plan for anonymous proposal reviews, starting with the HST Cycle 26 Delta TAC. The WG was asked to consider the implications of adopting a fully anonymous proposal review system, including identifying the appropriate process and modifications to the current proposal format. We were asked to offer instructions and guidelines to the community and to reviewers on how to write anonymous proposals, and how to review them in the suggested system. Lastly, the WG was tasked with engaging in a dialog with the community to solicit input, and mitigate concerns.

This report is provided in response to that charge, and is the outcome of our work in understanding what has been done in other anonymous reviews, and our effort to integrate community feedback into a suggested process. In an accompanying website, <https://outerspace.stsci.edu/display/APRWG>, we provide more detail on the literature and websites we used in our review, and provide some notes from our discussions. It should be noted that while our recommendations were explicitly centered on the HST Cycle 26 process, we agree that essentially the same process could be implemented in the first cycle of JWST, with the benefit of the additional launch delay allowing some time to evaluate the successes and failures of the proposed system on the HST reviews.

#### 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.

## Community Feedback and Discussion

Community feedback was sought through email request during the period February 19 to March 7, 2018. A draft of the WG's FAQs and proposers and reviewers guidelines were made available as part of the feedback solicitation. Approximately 60 responses were received during the request time period. A series of roundtables were organized at STScl to get local feedback, as well. The group received some very positive feedback, as well as some responses that raised a number of interesting questions. Among those questions was how one might devise equitable ways for ensuring that the proposing team has the requisite technical expertise, how the team's scientific productivity could be understood, and how one would judge the potential for completing the proposed analysis and publish the HST results. Additionally, questions were raised about how proposers should refer to observational or theoretical ancillary datasets, research results, or tools that may be key to the proposal, but could also serve as clear identifiers. How conflicts of interest would be addressed under full anonymization also raised some concerns. Finally, several respondents emphasized that the community will require time to adapt to any revised proposal guidelines.

Options other than those outlined in the WGAPR draft guidelines were also offered by members of the community and were considered by this working group. In one, a semi-randomized selection process was suggested in which the best and worst quintiles (or a similar range) are identified by graders, the best would be accepted, and the worst rejected. Those proposals "in the middle" would then be selected randomly until the time allocation was filled. This proposal reflects a sentiment in the community that at some level proposal selections are essentially

random, an erroneous interpretation of the variation due to the subjectivity of reviewers. This concept also raised a number of concerns for the WG, as factors deemed important to the peer review such as, “is the science is worth doing?” or “is the resource request reasonable?” would become irrelevant in the selection of proposals. We were also of the opinion that such a process would undermine the panel’s ability to provide valuable review feedback. The WG concluded that such a process would have long-term detrimental consequences on the overall success of the mission and on the legacy value of the HST data.

A two-step process in which the proposals would be evaluated anonymously on the science merit alone in the first step, and then evaluated on the team’s expertise (a second step where anonymity would be waived) was also suggested by the community. This model offered a number of advantages and served as a backbone to the recommendations made in the remainder of this report.

## Recommendations

Based on the available literature, feedback from the community, and the discussions of the Working Group, it is our recommendation that the Institute move toward a dual-anonymous proposal process beginning with Cycle 26 HST in late 2018. We understand that a fully anonymous process requires active participation from community, and that there is notable apprehension as to what the effect of anonymizing will do to the scientific productivity of the observatory. We therefore recommend a phased approach, in which most of review is done anonymously and a sensibility check done at the very end of the review, as described below.

### Informing the HST Community

The HST community will need guidance on how to prepare their proposals to ensure they are sufficiently made anonymous. The WG has prepared an initial FAQ to help understand our motivations for an anonymous process, and issues that such a process helps to mitigate. We have also made some initial guidelines with specific instructions to proposers on what to do and what not to do in preparing their proposals, and with instructions to panelists and panel chairs on preserving the anonymity in the review. The Institute should further curate the FAQ and guidelines, and make them available with the Cycle 26 Call for Proposals.

The WG also recommends the Institute revise the orientation material provided to the panel and panel chairs, with a goal of making them more focused on the issues relevant to the review of the current cycle, with less focus on information or statistics of previous cycles.

### Conceptual Framework of the Dual-Anonymous Process

The WG recommends the following for this dual-anonymous process. Proposals will be written in a largely anonymous fashion, where the sections which include the scientific justification and

technical description will be anonymized following the recommended proposer guidelines. Proposers are encouraged to write in a passive voice when discussing their work as they would for any other relevant work in the field, e.g., using phrases such as “As Doe et al. have shown...” as opposed to “As we have shown in Doe et al...”. Proposers may not include statements that directly identify individuals as participants in the proposal team, including identified self-references.

Additionally, proposers will provide a brief (no more than a page) section on team expertise and background. This additional free-form section will describe the requisite areas of expertise crucial to the success of the program, and the roles of team members in those areas. As such, this section is not expected to be made anonymous. This section will, however, be withheld from panelists and chairs throughout most of the in-person review.

Once the proposal ranking is complete, as a final sensibility check, panelists will be given the names of the investigators and their expertise sections, and may flag to disqualify proposals they deem as insufficiently poised to carry out what they’ve proposed. Potential disqualifications should be decided by a consensus (or at least a majority), and flagged in the final panel report to the Director. The panel should not adjust the scientific ranking of the proposals.

## Recommended Modifications to the Phase I Process

The WG recommends a mostly dual-anonymous review of the Phase I submissions, with the following modifications.

1. Proposals should be crafted to be anonymous following the proposer guidelines provided. Proposals that are egregiously non-compliant with the anonymizing guidelines, as identified in the preliminary grading (or triage stage), may be flagged for removal from further review.
2. Before the panels/TAC meets, reviewers should be briefed on the expectations for discussing the anonymized proposals and receive training in avoiding unconscious bias.
3. The panels should indicate which proposals they collectively found were not sufficiently made anonymous, according to the proposal guidelines. Those proposals will be flagged in the TAC recommendations to the Director.
4. It is expected that the panels will include specific feedback in their comments to investigators if their proposal were not sufficiently made anonymous.
5. Proposers will be required to submit a Team Expertise and Background section with their Phase I submission. This section will not be anonymous and would be available to the reviewers as part of a final sensibility check. The format should be free-form, and could 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 be a description of an analysis plan, including a description of the effort and the roles of the investigators (if known). It need not be the level of detail of a program management plan. An example of such a Team Expertise and Background section should be provided.

6. Once the review is nearly complete, and the proposals are scientifically ranked to well beyond the panel's nominal orbit allocation, the panel should then be given the list of investigators and the Team Expertise and Background section, for all programs above the panel's allocation line. In this last sensibility check, panelists will be permitted to flag proposals for disqualification if the motivation for doing so is well-justified and agreed to by a consensus. The rationale for flagging a proposal should be clear, compelling deficiencies in the expertise required to see through the goals of the proposal.
7. Levelers, persons charged with monitoring the panel review process in person from start to finish, should be used to insure the guidelines for proposal reviews are followed, especially during the final sensibility check. Levelers should be responsible for ensuring the discussions of the panel remain focused on the scientific merit of the proposal. In this way the role of the Levelers is essentially the same as is used in NASA reviews.

## Recommendations for Assessment

The Institute should continue to monitor the productivity, diversity and gender balance, and overall scientific impact of awarded programs, utilizing its resources in the STScI library and archive. It may take a few cycles for the impact of these changes to be realized in publication rates, and even longer for citation rates. However, these figures should be shared periodically with the STUC. Additional metrics such as the number of new proposers and the success rates of junior and senior PIs may be easily compared to past performance, and could provide a shorter-term indication of improvement.

We'd like to thank the STScI Chief Librarian, Jenny Novacescu, for compiling a large number of articles on dual-anonymous peer reviews that have been highly useful in our discussions. We also thank Jessica Kirk (U. Colorado) for her contributions to the discussion and to this report.