Thoughts from the AGU SPA Fellows Committee

Alexa Halford, Angeline G. Burrell,
Endawoke Yizengaw, Volker
Bothmer,
Brett A. Carter, John Raymond, Astrid
Maute, Marilia Samara, Naomi Maruyama,
Please feel free to add your name,
I didn't want to assume anyone wanted to be
on
here but it would be great to have the entire committees as co-
authors.

March 2021

Contents

1 Introduction from Dr. Halford, the previous Chair

I served on the AGU Space Physics and Aeronomy (SPA) Fellows committee from 2017 - 2020, chairing it in 2019 and 2020. At first, like many, I was not confident that I completely understood the award process. Today, I recognize that each section and committee work a bit differently and that the award criteria are interpreted in various ways each year as the committee members change. I believe this and the fact that the definitions and interpretations of the award criteria are not shared with AGU members are key reasons why it is continually confusing why some nomination packages succeed while others do not. Through this communication, my co-authors and I aim to shed some light on how our committee approached this task, increase the transparency of the

process, and detail the steps we took to mitigate and remove systemic biases. We also hope that future committees will continue to improve transparency. Furthermore, by increasing transparency, we hope to encourage community members to submit nomination packages.

One thing that remains consistent in each section's committee from year to year is the solemnity that each member brings to the table. All the committee members show the highest respect for each nominee's excellent contribution to the field. However, each committee does, and must, work differently. Factors contributing to this include the number of packages, which can vary significantly from year to year, and the geographic distribution of the committee members. The SPA section typically receives between 20 - 30 packages to evaluate within a month, which falls roughly in the middle compared to other AGU sections. This time constraint means that each SPA package receives, on average, about 12 minutes of group discussion. This time does not include the time invested by individual committee members, who (during my leadership) read over all the individual packages and delve in-depth into X number of packages. Working with such dedicated committees striving to bestow recognition on our incredibly deserving colleagues has been an honor.

Before we get into the details, I want to applaud and acknowledge all of our committee's work. Our committee comprised 12 individuals from across the world and the SPA disciplines. They were asked to do a substantial amount of work in a concise amount of time. They did so with complete professionalism and diligence and without complaint. Committee members made great efforts to attend meetings while at conferences and on travel. Many went above and beyond by making meetings at times well outside of reasonable working hours when they would have otherwise been asleep or managing other necessary aspects of being a living human. As the chair, I am incredibly thankful for their dedication to making this work, not least because, for many, the hours they dedicated to working on this committee came from their personal time.

2 Criteria for Selection:

AGU has laid out three criteria for nominating a fellow AGU member (https://www.agu.org/Honor-and-Recognize/Honors/Union-Fellows [1]):

1. Breakthrough and/or discovery,
2. Innovation in disciplinary science, cross-disciplinary science, instrument development, or methods development, and
3. Sustained scientific impact.

The SPA committee did not prioritize one category over another, nor did we systematically consider whether or not a candidate met the criteria in more than one category. As a result, these criteria can be subjective. Before viewing the nomination packages, our committee discussed the evaluation criteria to have a common interpretation. The interpretation of the evaluation criteria can also be subjective and a potential source of confusion for nominators. The interpretation may change from year to year and from section to section. One example is the h-index, which is listed as an optional metric to include in the nomination package [1]. Other optional metrics that one can include are not listed, such as the number of successful Ph.D. students or the number of instruments built and flown. By listing the h-index as an option on the AGU website, its perceived value as a shortcut metric is elevated above other metrics. Well-known biases are associated with the h-index,

including biases that affect women, minorities, and fields or sub-fields that publish at different rates [2, 3, 4, 5, 6, 7]. Given the well-documented biases of the h-index, which does not consider the quality of the research, we strongly recommend excluding it as an evaluation criterion, even an optional one.

2.1 Defining and interpreting the Evaluation Criteria

Our committee decided that there should not be any predetermined order or weight to the itemized definitions or criteria. Each of these evaluation criteria are defined in detail below as provided by AGU.

Breakthrough or Discovery: An idea that once accepted, allows others to frame ideas or approach problems differently and more effectively than before.

Innovation in disciplinary science, cross-disciplinary science, instrument development, or methods development:

- Enabling collaborations across many sub-fields.
- Development of new instruments that have been successful in the field and lead to new* understandings.
- Development of new* methods that other scientists have adopted and have led to new* understandings within the field.
- Produced a data product or a method that is used on a routine basis even if not correctly cited. (Has an open data/code policy and has become so routine, people have forgotten that this is either produced by someone or was not a standard product previously.)

*New: something that deviates enough from 'standard understandings' in any one field in the presented form, even if the process to arrive at 'new' happened through a series of gradual improvements or advancements.

Sustained scientific impact:

- Something that has changed the way other scientists approach a problem, perhaps on a smaller scope but cumulatively changes people's perceptions over time.
- Enabled long-lasting collaborations leading to significant impact within the field.
- Mentor a significant number of collaborators/scientists/students, enabling their development as researchers.
- Produced continued excellent research over the course of their career.

The SPA committee definitions and interpretations are still general, and perhaps not fully inclusive. We used this to establish a *lingua franca* within the committee, aiding discussions throughout the evaluation process.

2.2 Evaluation Process:

Our committee took the evaluation process very seriously. We also were very aware that the SPA section has previously failed to recognize all portions of our community justly (e.g., gender, race, or ethnicity) [8]. We were also keenly aware that we held similar implicit biases as members of our own cultures and research sub-fields. The first step the chair took, with the help of our SPA president at the time, was to attempt to mitigate the impact of our implicit biases by constructing a balanced committee. For the last few years, our SPA Fellows nomination committee comprised a nearly equal number of men and women and nearly equal representation from the solar, interplanetary, magnetosphere, and ionosphere/atmosphere communities (the major sub-fields within SPA) and representation from across the globe and career levels. Dr. Halford was the most early-career committee member (currently nine years post Ph.D.), with others spanning among the most senior ranks of our field. This committee construction aimed to gather people with contrasting implicit biases so that the impact, on average, could be mitigated. While our rankings showed that we still held implicit bias for our sub-fields, these were mitigated by having our diverse committee resulting in an equal distribution of each sub-fields within the rankings. For example, if we had had a persistent magnetospheric bias in our committee members, it would have been likely that more magnetospheric nominations would have been put forward.

The broad time zone difference between our committee members also meant we needed to consider the best times and methods for the meeting. We took two approaches: staggering meeting times and maintaining an online repository. Each week we had two meetings, one that was not at obscene hours for those in Europe/Africa and another that was not at obscene hours for Australia/Asia. In addition, our shared online repository was accessible and editable by all members and allowed all committee members to access the notes made by others about each nomination package. The two steps we took (thoughtful committee construction and moderated committee interactions) laid a solid foundation for the success of our meetings. Without these two steps, we would have likely still put forward deserving nominees. However, as demonstrated by the prior selections, these nominees would not have represented our community.

During the committee's first meeting, we discussed the different types of biases we may each hold, so we reminded ourselves that we should be conscious of them throughout the rest of the process. Below is the list of potential biases we identified and attempted to mitigate through a balanced committee and open discussion.

- Gender
- Nationality
- Race/Ethnicity
- Career level (retired/senior/expert vs mid or even mid/expert/senior)
- Extrovert vs Introvert (impacting who is seen, heard and remembered)
- A country or institution's socioeconomic status (e.g. the ability and opportunity to network in person)
- Large Mission participation vs smaller projects such as CubeSats, rockets, balloons etc.

- Experimentalist vs theorist vs observationalist
- Dependence on intrinsically biased, short-cut metrics
- Bias towards our own sub-fields
- Publication/collaboration environment
- The Matthew Effect (credit being attributed to the most well-known name, not the person who necessarily had the ideas, did most of the work or is listed as first author [9].
- The Matthew/Matilda affect (where men tend to get the credit more so than women who did just as much or more of the work) [10, 11].
- Work in "up-stream" fields. For example, much of solar physics impacts the other sub-fields, but the ionosphere does not impact the sun.
- Work in a traditional academic environment
- Multidisciplinary work
- Number of awards received.

We took a broad view and discussed how these biases might affect our perspective on the nominee's impact on the field. Sometimes these biases lead to positive or negative impacts and perceptions. For instance, we had multiple discussions on how work done by a mentee should be considered in a nomination package for their mentor. This is related to the Matthew effect [9]. Were the nominees being given the credit that the mentee deserved (especially when the package presented the work done by the mentee as a breakthrough or discovery by the mentor), or should they be getting credit for supporting and collaborating with the mentee (an excellent example of sustained scientific impact)? For cases like this, how a nomination package presented the work significantly impacted the committee's perception.

Many of the identified biases were found to affect shortcut metrics of a nominee, such as the h-index, [4]. For example, the types of projects/platforms/work environments can significantly impact the number of papers a person is likely to write or be a co-author. If a person works within a larger collaborative group, they are likely to be on more papers with a large number of co-authors [4]. Specifically within space physics, as in many fields, the number of co-authors was found to be correlated to the number of citations [12]. Another factor that can impact the number of co-authors is visibility within the field which can lead to more extensive and more diverse collaborations [13]. For example, are the nominees able to attend conferences regularly, and are they invited to speak and give presentations [14]? The number of papers and citations were found to bias the perceived prestige of the project and, thus, the nominee associated with that project instead of the impact and quality of the work. In addition, shortcut metrics such as the h-index was found to move the discussion away from the substance of the publications and did not leave room for sometimes vitally important data sets like geomagnetic indices, which are frequently improperly referenced in publications. We discussed data sets such as these and other tools that are now considered well-understood standards and "owned by the community" [6] for each nominee's package.

With the recognition of the impact of a nominee's work in the field, the identified biases can have an impact on a nominee's package. We took an additional step to mitigate these biases. As a committee, we worked towards creating a safe environment where any given member felt empowered to speak up when they observed the influence of biases on the discussion. This was accomplished by first addressing the issue of bias via email. AGU also addresses these issues in the orientation for the committees. We further discussed and were open with each other about our own biases during the first meeting. As the chair, I asked a few of the committee members I was close with to make sure to call me out on biases. This was done to help show that it is okay to be called out and that we are participating in helping each other to ensure that we are putting forward the most deserving scientists from our field. At least once during each meeting, we asked if anyone had noticed any biases during the discussions.

All committee members read all nomination packages. Many committee members (if not all) also read the papers referenced in the supporting letters and the bibliography. The materials in the nomination packages help provide evidence for the nomination citation and subsequent claims made within the nomination package supporting the evaluation criteria. Some initially broke the packages into three groups, top, middle, and bottom, to help focus discussions. Much of the discussions revolved around what evidence was presented, what was omitted, and if the nomination and supporting letters were consistent with the short citation, CV, and the selected bibliography.

Each meeting for the committee was scheduled to be two hours long, allowing for approximately a 12-minute discussion per package. The discussions were timed to ensure each package had a similar amount of discussion time. If more discussion was needed for a particular package, this was noted to return to it if time allowed. Committee members presented the packages and led discussions about what achievements were described and had evidence related to the three previously outlined criteria. If members could not attend the meeting or felt more comfortable providing written comments, they contributed their notes and comments to the summary for the nominee through the shared drive.

During the final set of meetings, we discussed the ordering of the nomination packages. We considered multiple ranking strategies, including the mean rank, median rank, the number of 1's, 1's and 2's, etc., each nominee received, and more. We found that with few exceptions, the ranking of the nominee did not change much (typically no more than a shift of 1 - 3 positions) when using any of the given methods. This gave us confidence in our choices put forward to the union committee and their final order. In instances when the ranking changed significantly, or if the shift occurred at a critical boundary (e.g., changed who would be put forward to the Union Committee), we considered the deviation between the rankings. We discussed the reasons behind any scores that significantly differed from the majority opinion. We also took the time to check our potential biases. Given the distribution of submitted nomination packages, we found a relatively even distribution of sub-fields, gender, and other underrepresented groups. We feel confident that through a diverse committee and discussions about potential biases, we have sufficiently mitigated our biases and put forth the most deserving nominees.

The top four candidates are typically unanimously supported by the committee. The recent committees have been satisfied that this group is representative of the best within all sub-fields of our community. The most contentious packages were those whose nominated work undoubtedly contributed to our field but did not address the connection between their work and the SPA sub-fields. It is sometimes unclear what the best route is to take with these nominations, which are also

usually dual submissions with another field such as planetary or Atmospheric and Space Electricity.

3 Recommendations Pertaining to the Program Process:

At the end of the committee's work, we tried to reflect on and identify any other biases that may have affected our discussions and rankings and add them to the list for the following year. For example, after my first year as chair, we identified three more potential biases that were added to our list. The first new bias identified was whether the science was a part of the Solar (SH) community. The data products and scientific results from this sub-field are frequently utilized by the magnetospheric and ionospheric/atmospheric communities. However, solar scientists are frequently unaware of the work performed within the magnetospheric and ionospheric/atmospheric communities. The impact of this physical reality was seen both in the applicability of a topic to interdisciplinary science and in the likelihood of journal articles obtaining a high number of citations. During committee discussions, we determined that some aspects of this bias are not actively harmful. Each SPA sub-field has a different scope, and unlike the solar community, the magnetospheric and ionospheric/atmospheric results may be perceived as having a more immediate impact on society. This could interact with the experimentalist/theorist bias. Scientific advances in these sub-fields may be unconsciously interpreted as being more applied science and thus less worthy of being considered a discovery or breakthrough. Although the committees have been unable to determine the best way to address this bias, it was identified and discussed. Second is the number of other fellowships or awards won by a nominee. This, again, was not consistently perceived as good or bad. Some committee members perceived a large number of awards as a reliable indicator of quality science, as it showed how distinguished a person was. Others perceived the presentation of awards negatively, as they did not consider it a reliable short-cut metric for excellence, and it took up space that could have been spent directly discussing the scientific impact made by the nominee. Still, others perceived this negatively because they felt we should acknowledge those who did not already have many awards but were still highly deserving.

We found it critically helpful to discuss biases and the evaluation criteria we would use. We also found it beneficial to have these discussions before reading and ranking the nominations. It provided a moment for everyone to check their thought process before verbalizing an opinion and identify which biases had affected their interpretation of a package.

AGU's nomination package and software play into at least one implicit bias we all have. Committee members use this site first to gain access to the nomination packages. The first tab shows up, and the first bit of information when you download the package pertains to the nominator. It should not matter who the nominator is, and putting it up front gives the impression that this is more important than the nominee, thus promoting and perpetuating the idea that science is still at the whim of the "good ol' boys club". Putting the nominee information up front would help mitigate the Matthew bias and return the emphasis where it belongs: on the nominee's skills and accomplishments.

4 Conclusion

The Fellows honor is considered the highest honor AGU can bestow. Because of this, we believe it is paramount that the evaluation criteria reflect the values of our community. The vast majority of nomination packages we receive undeniably deserve high praise for the nominee's work and

commitment to the AGU community. However, in the past, some very important values have been overlooked. These typically fall under the "sustained scientific impact" section of the AGU Honours nominating criteria and include: the impact of service and sustained support activities, such as data curation, which enables countless others to lead breakthroughs and discovery or perform cross-disciplinary work. There is also a long history of ignoring the breakthroughs and contributions made by individuals from underrepresented groups and those with less political power. This includes women (~12% of current SPA fellows) and racial/ethnic minorities (< 12% of current SPA fellows), among others. These biases against marginalized groups and institutions can be mitigated by avoiding heavy-weighting metrics such as h-index and past awards [6, 5]. For example, within the SPA community, we have had years where zero women were nominated. This has led to many biases concerning who has become, or more accurately, not become, a Fellow in the past. This has improved in recent years thanks to the efforts of the Nominating Task Force [8] and this committee and AGU's efforts to acknowledge and mitigate implicit biases. However, we must continue to be vigilant and work towards ensuring we recognize all those who are deserving of becoming an AGU fellow. We encourage the AGU community, Union section, and AGU leadership to reflect on this as we continue to consider biases within our fields. Furthermore, we work toward ensuring that colleagues who have been forgotten because of the "invisible" work they do are honored according to their contributions.

References

[1] AGU. Union fellows. *website*, 2021.

[2] Kristoffer Rørstad and Dag W. Aksnes. Publication rate expressed by age, gender and academic position – a large-scale analysis of norwegian academic staff. *Journal of Informetrics*, 9(2):317–333, 2015.

[3] Elissa Z. Cameron, Angela M. White, and Meeghan E. Gray. Solving the Productivity and Impact Puzzle: Do Men Outperform Women, or are Metrics Biased? *BioScience*, 66(3):245–252, 01 2016.

[4] Iman Tahamtan, Askar Safipour Afshar, and Khadijeh Ahamdzadeh. Factors affecting number of citations: a comprehensive review of the literature. *Scientometrics*, 107(3):1195–1225, 2016.

[5] L. Leydesdorff, L. Bornmann, and T. Opthof. h_α : the scientist as chimpanzee or bonobo. *Scientometrics*, 118:1163–1166, 2019.

[6] Colin A. Chapman, Júlio César Bicca-Marques, Sébastien Calvignac-Spencer, Pengfei Fan, Peter J. Fashing, Jan Gogarten, Songtao Guo, Claire A. Hemingway, Fabian Leendertz, Baoguo Li, Ikki Matsuda, Rong Hou, Juan Carlos Serio-Silva, and Nils Chr. Stenseth. Games academics play and their consequences: how authorship, h-index and journal impact factors are shaping the future of academia. *Proceedings of the Royal Society B: Biological Sciences*, 286(1916):20192047, 2019.

[7] T. Pico, P. Bierman, K. Doyle, and S. Richardson. First authorship gender gap in the geosciences. *Earth and Space Science*, 7(8):e2020EA001203, 2020. e2020EA001203 2020EA001203.

[8] A. N. Jaynes, E. A. MacDonald, and A. M. Keesee. Equal representation in scientific honors starts with nominations. *Eos*, 2019.

[9] Robert K. Merton. The Matthew effect in science. *Science*, 159(3810):56–63, 1968.

[10] Anne E Lincoln, Stephanie Pincus, Janet Bandows Koster, and Phoebe S Leboy. The matilda effect in science: Awards and prizes in the us, 1990s and 2000s. *Social studies of science*, 42(2):307–320, 2012.

[11] Margaret W. Rossiter. The Matthew Matilda effect in science. *Social Studies of Science*, 23(2):325–341, 1993.

[12] Mark B. Moldwin and Michael W. Liemohn. High-citation papers in space physics: Examination of gender, country, and paper characteristics. *Journal of Geophysical Research: Space Physics*, 123(4):2557–2565, 2018.

[13] Nader Ale Ebrahim, Hadi Salehi, Mohamed Amin Embi, Farid Habibi, Hossein Gholizadeh, and Seyed Mohammad Motahar. Visibility and citation impact. *International Education Studies*, 7(4):120–125, 2014.

[14] Heather L Ford, Cameron Brick, Karine Blaufuss, and Petra S Dekens. Gender inequity in speaking opportunities at the American Geophysical Union Fall Meeting. *Nature communications*, 9(1):1–6, 2018.