STScI | SPACE TELESCOPE SCIENCE INSTITUTE

EXPANDING THE FRONTIERS OF SPACE ASTRONOMY

Scaling the TAC

Laura Watkins

Science Policy Division

December 2, 2025

Workload & Oversubscription Rate

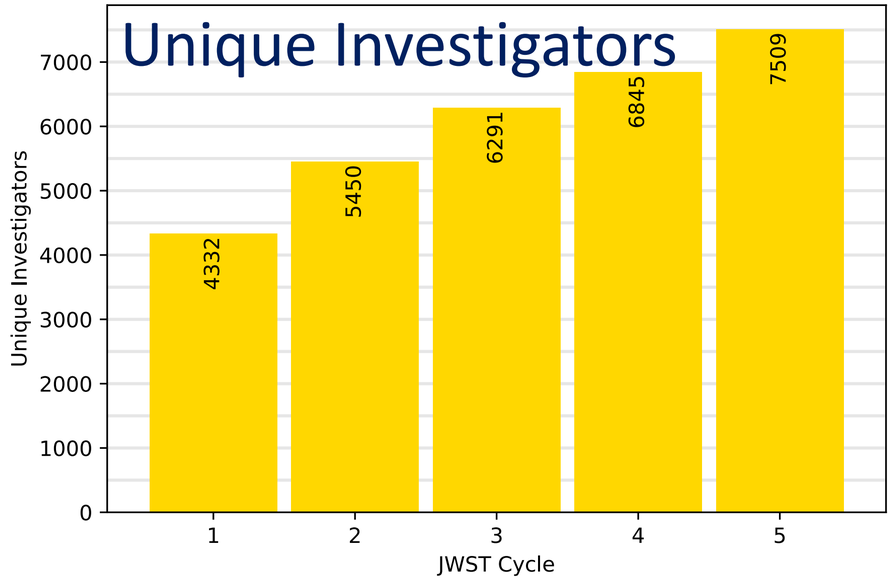
- Workload for reviewers and oversubscription are separate though related issues
  - **Oversubscription:** amount of time requested vs time available
  - **Workload:** number of reviewers vs number of proposals

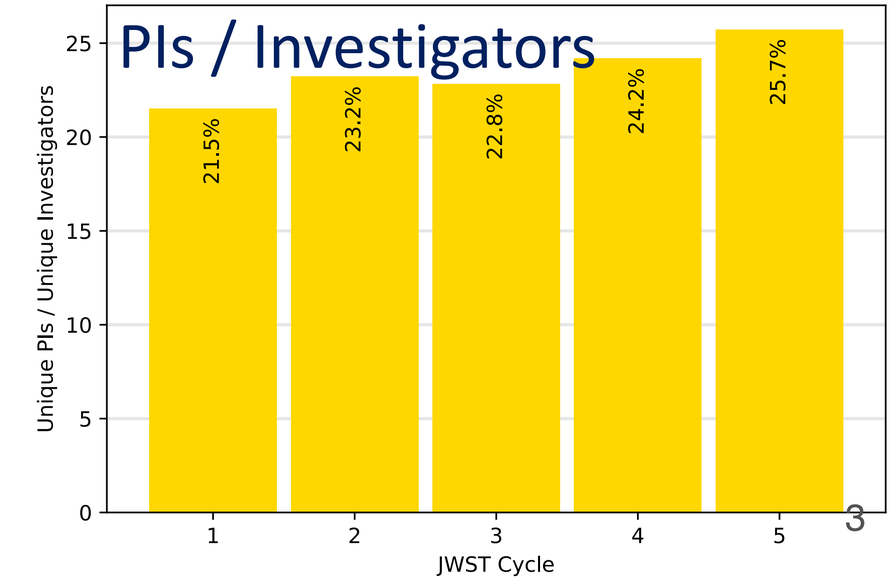
- Should we control the oversubscription?
- How well does the current TAC model scale under reasonable workloads?

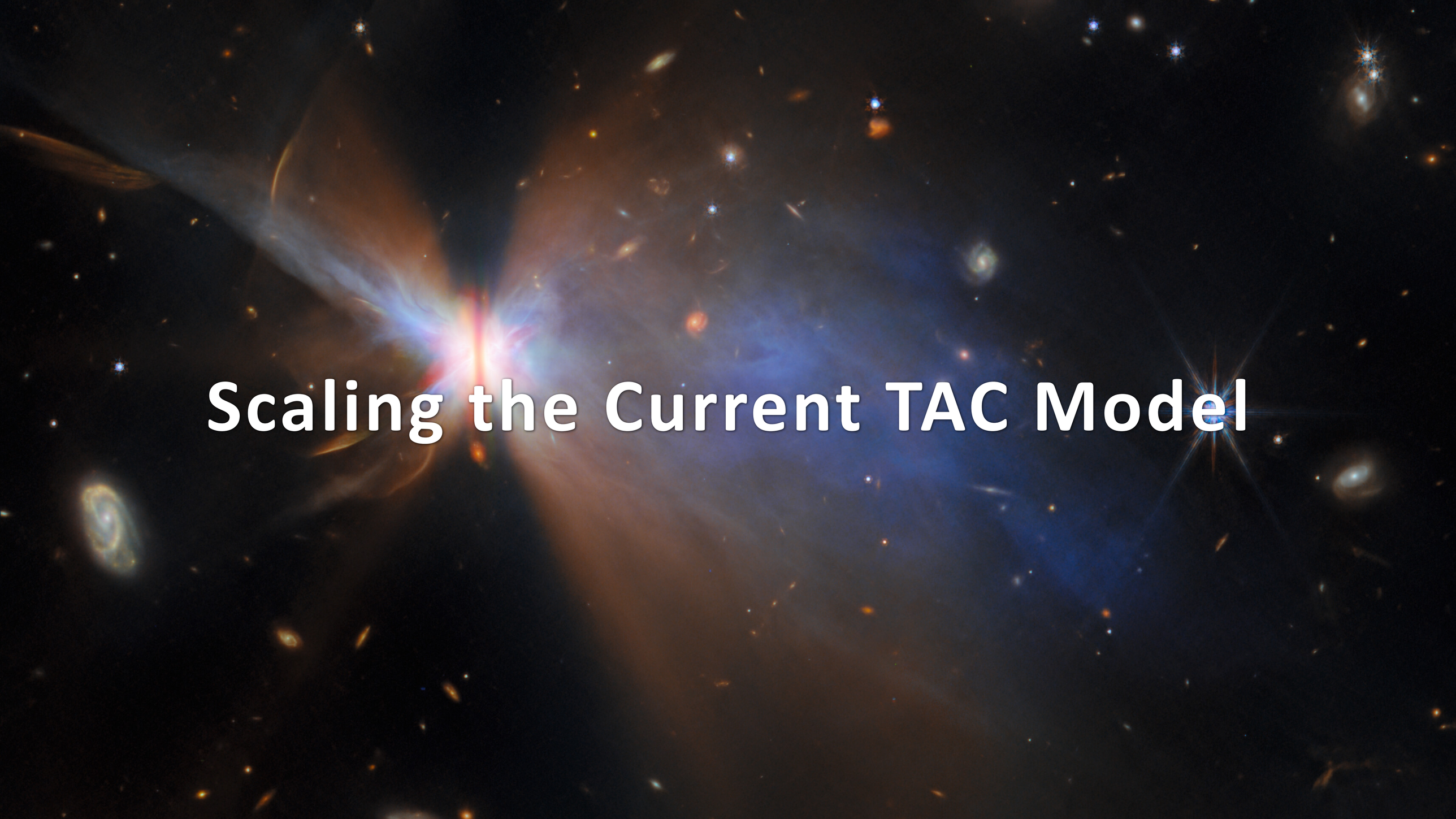
Should we control the oversubscription? (No.)

- **JWST Community is increasing each Cycle.**

- **Capping proposals:**
  - **Limits science ideas.**
  - **Inhibits innovation.**
  - **Artificially reduces apparent demand for the observatory.**







Scaling the Current TAC Model

Telescope Allocation Committee (TAC) - Current Process

- 8 science categories - grouped in Galactic + Extragalactic
- **Executive Committee - workload typically 35-45 proposals:**
  - 2 for Cycle 5 (1 Galactic, 1 Extragalactic)
  - Large GO, Treasury GO, AR Legacy, Pure Parallel
- **Discussion Panels - workload typically 50-60 proposals:**
  - 17 for Cycle 5, subdivided by science category
  - Small GO, Medium GO, ToO, Survey
- **External Panels - workload typically 20-25 proposals:**
  - 8 for Cycle 5, subdivided by science category
  - Very Small GO, AR

Reviewers & Invitations for Cycle 5

- **961** invitations for **550** reviewers
- External panelists accept at a higher rate than Discussion panelists
- Consistent: Cycle 4 was 46% for Discussion, 66% for External, 57% overall

Cycle 5 TAC				
	Acceptances	**Declines**	**Total Invites**	**Success Rate**
Executive Committees	37	24	61	61%
Discussion Panels	190	227	417	46%
External Panels	323	160	483	67%
Total	550	411	961	57%
ESA	199	148	347	57%
CSA	13	29	42	31%

How well does the TAC scale?

- **We don’t want workloads to increase so…**
- **More proposals → more reviewers!**
  - There’s a limit to which this is a feasible solution without making other changes.
  - Need to consider difficulty of recruitment/community agreeing to serve (not asking too many folks each time, keeping workloads manageable, making it an enjoyable experience)
  - Need to consider STScI resources/how many panels and panelists we can support.

How well does the current model scale?

- **Knobs we can turn under the current model:**
  - Which size/class/type proposals go to EC, Discussion, External.
  - Number of proposals per panel
  - Number of reviewers per panel
  - Number of prelim/external grades per proposal
    - *(we have already dropped from 6 prelim grades to 5 for discussion panels in Cycle 5)*
  - Triage level of discussion panels

Current TAC: Discussion Panels & Executive Committee

- **Discussion Panels & Executive Committee:**
  - 1-3 panels per science category depending on proposal load, 2 ECs
  - R reviewers per panel
    - R needs to be big enough to manage conflicts and spread workload but low enough to keep discussion productive (generally R=8-12)
  - Preliminary phase: panelists read and grade a subset of proposals P
  - Triage: remove the lower T%, remaining proposals are discussed
    - Discussion list needs to be large enough for good depth of quality and perspective of panel, but small enough to be manageable in time available and to avoid panel frustration if fraction of proposals recommend is too small
  - Extra reading: panelists won't have read everything on the Discussion List, read extra E proposals
  - Discussion meeting

Current TAC: External Panels

- **External Panels:**
  - 1 panel per science category regardless of proposal load
  - R reviewers per panel
    - R needs to be big enough to manage conflicts and spread workload, recruit as many panelists as needed to meet projected workload
  - Panelists read and grade a subset of proposals P

Intentionally Ridiculous Scenario

- **What if we got 5000 proposals in a cycle?**
  - We **don't expect** this many in a cycle! But if we can work up a feasible TAC model in this scenario, we can do it for more modest proposal numbers too.
- Under a Cycle 4 (5) like TAC model, we would need:
  - Discussion: 386 (328) reviewers in 36 (28) panels
  - External: 566 reviewers in 8 panels
  - EC: 72 (56) reviewers in 4 panels
  - **Total: 1024 (950) reviewers**

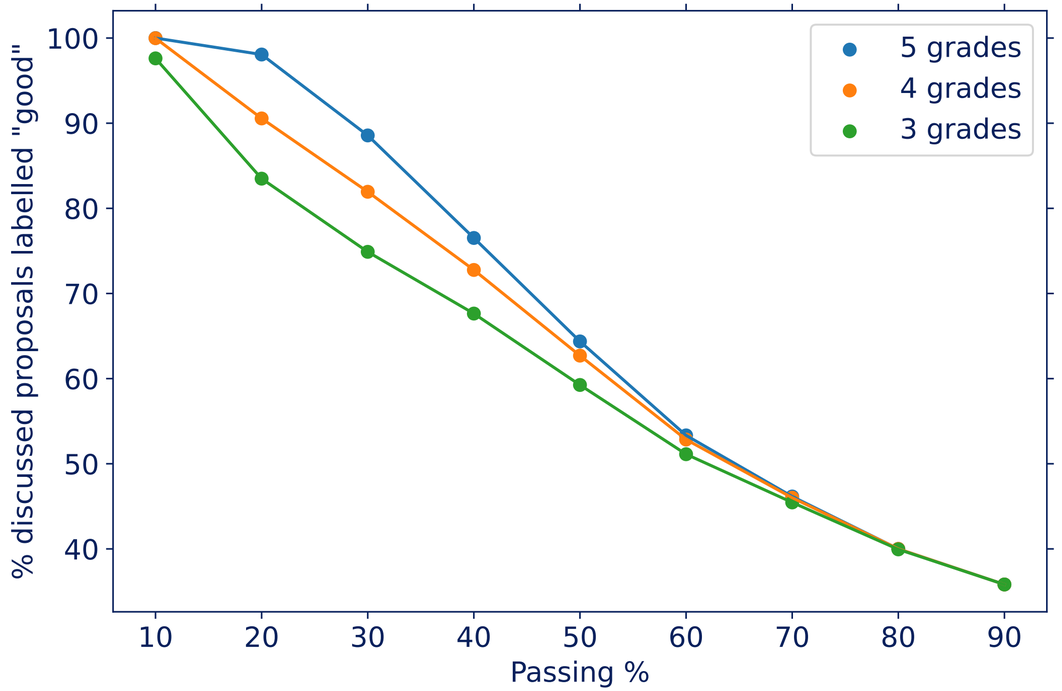
Serious Model for the Ridiculous Scenario

Not the only solution!
But shows it's possible.

- Assuming a total of **5000 proposals** subdivided by type and science category as Cycle 4:
  - Target workload: **30 proposals per reviewer** for External / Prelim, full Discussion workload ~55-60 proposals when accounting for extra reading
    - Currently 20-25 for External
    - Currently 40-45 for Prelim (resulting in 50-60 total) for Discussion
    - Currently 20-25 for Prelim (resulting in 35-45 total) for EC
  - Grades per proposal: **3 grades for External and Prelim** (all available for Discussion)
    - Currently 5 for External and Prelim (was 6 for Prelim in Cycle 4)
    - Tests show that 80% of untriaged proposals remain unchanged going from 5 to 3 reviewers.
  - Triage: **70% for Discussion and EC** (not applicable for External)
    - Currently ~50% but varies panel to panel

Reducing the number of grades

- **How many “good” proposals pass triage with reduced grades?**
- **Define good:** HST success rate ~1:6. Panels set do-not-observe line at 2x allocation line **or lower**.
  - Implies **at least** ~1:3 proposals (33%) are “good” and worthy of observing.
- **Test with real JWST panel:** 75 proposals, 10 reviewers, 6 preliminary grades per proposal.
- Re-ran rankings using G=5/4/3 grades, 1000 runs for each G model. Triaged each run using passing cuts of 10%-90%.
- X-axis: passing cut (% of panel proposals sent for discussion, 100-X=triaged %)
- Y-axis: mean percentage of the passing/ discussion proposals that are “good” proposals

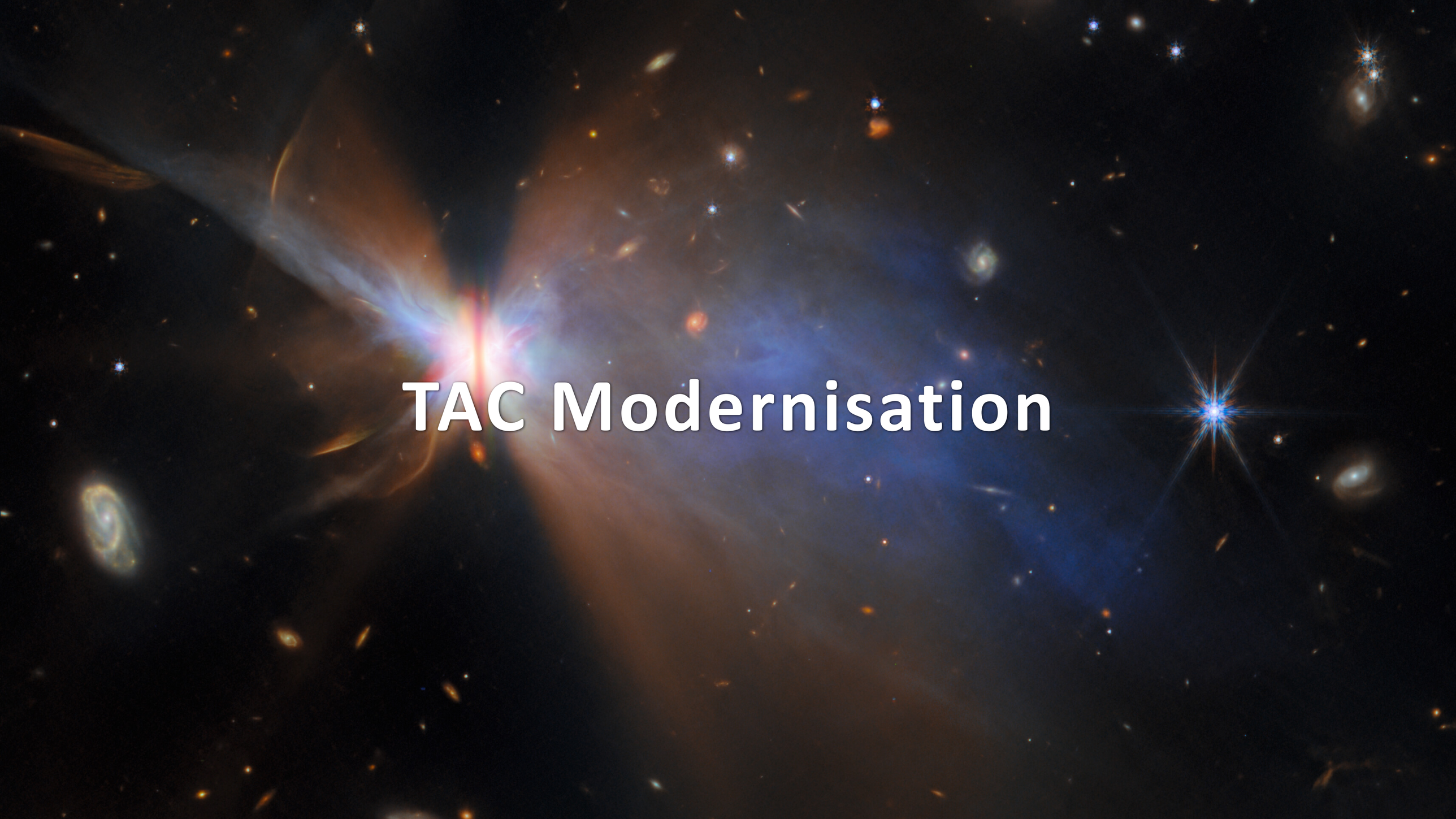


- For 3 preliminary grades, and a passing fraction of 30% (triage 70%):
  - ~75% of discussed proposals are “good”
  - Only ~10% of the total panel expected to be recommended (or 33% of the discussed proposals)
  - Panel discusses 3x as many proposals as can be accepted
  - Panel discusses 2.25x as many “good” proposals as can be accepted

Serious Model for the Ridiculous Scenario

Not the only solution!
But shows it's possible.

- Regroup science categories:
  - **Planetary:** Solar System; Exoplanet Atmosphere; Planetary Systems
  - **Galactic:** Stars and Stellar Pops; Dust, Gas and the ISM
  - **Extragalactic:** Nearby Galaxies, High-z Galaxies, SMBH & AGN
- Move Smalls from Discussion to External, and make Discussion panels broader
  - **External: Very Small + Small + Archival**
    - need 392 reviewers in 8 panels corresponding to 8 science categories
  - **Discussion: Medium + ToO + Survey**
    - Planetary(2), Galactic(2), Extragalactic (3) → need 88 reviewers in 7 panels
  - **EC: Large + Treasury + Pure Parallel + AR Legacy**
    - Planetary+Galactic(1), Extragalactic (1) → need 24 reviewers in 2 panels
- Total: **504 reviewers** (for comparison for Cycle 5 we have ~550 reviewers)


TAC Modernisation

Guiding Principles

Community TRUST in our TAC processes is fundamental.

Science is a HUMAN endeavour. Proposals should be evaluated by humans, not machines.

Modernising TAC processes & exploring AI tools

Community TRUST in our TAC processes is fundamental.

Proposals should be evaluated by humans, not machines.

- Examples of areas that will benefit from modernisation (AI or otherwise):
  - Reviewer recruitment — identifying possible reviewers and suggesting in which science area/ panel they would fit best
  - Providing input for assigning proposals to science areas and panels — could even be incorporated in APT as a suggestion to proposers
  - Providing input for assigning proposals to reviewers — including balance in degree of expertise overlap
  - Identifying problematic language in review comments — could be included in the review software for reviewers in real time
  - Page limit and formatting compliance checks
  - Identifying possible DAPR compliance issues for further investigation

Modernising TAC processes & exploring AI tools

Community TRUST in our TAC processes is fundamental.

Proposals should be evaluated by humans, not machines.

- Currently investigating where and how AI tools could be used to **support** TAC work:
  - This will take the time to do right. Targeting first use in Cycle 6, but will take a few cycles to roll out fully.
  - Carefully constructing the training inputs and thoroughly verifying the AI outputs will be our highest priority.
  - We will be clear with the community about where and how AI is used.
  - All materials used for training or review will be kept **confidential**.
  - This will be a **tool** to help humans do their work but won't replace the human care and effort—from reviewers and STScI staff— that underpins our TAC processes.
- **Also investigating non-AI solutions.**

Additional Info - Recruitment Stats

Current TAC: Reviewers & Invitations for Cycle 5

Executive Committees				
	Acceptances	Declines	Total Invites	Success Rate
Galactic	21	13	34	62%
Extragalactic	16	11	27	59%
Total	37	24	61	61%

Current TAC: Reviewers & Invitations for Cycle 5

Discussion Panels				
	Acceptances	Declines	Total Invites	Success Rate
Solar System	13	25	38	34%
Exoplanet Atmospheres	36	29	65	55%
Planetary Systems	20	24	44	45%
Stars & Pops	27	42	69	39%
Dust & Gas	17	25	42	40%
Galaxy Ecosystems	19	11	30	63%
Galaxy Frontiers	34	51	85	40%
SMBH	24	20	44	55%
Total	190	227	417	46%

Current TAC: Reviewers & Invitations for Cycle 5

External Panels				
	Acceptances	Declines	Total Invites	Success Rate
Solar System	27	22	49	55%
Exoplanet Atmospheres	37	5	42	88%
Planetary Systems	46	23	69	67%
Stars & Pops	52	42	94	55%
Dust & Gas	39	30	69	57%
Nearby Galaxies	40	12	52	77%
High-z Galaxies	38	11	49	78%
SMBH	44	15	59	75%
Total	323	160	483	67%