Time Allocation Strategies: Lessons from HST

Neill Reid (STScI)

TMT Forum

22 June, 2015

Context

- Every observatory aims to devise an equitable and efficient process for allocating observing time to it's user community
  - The process should aim to optimise the scientific return from the facility.
  - The process is defined by the institute running the observatory, but is only successful if supported by the community
- Most observatories employ a form of peer review to select proposals
  - HST convenes a committee comprising ~135-145 scientists drawn from the US & international community
    - ~1,000 – 1,100 proposals per cycle
    - Topical panels with individual resource allocations based on proposal pressure
    - ~60-90 proposals per panelist
  - ESO, Chandra, ALMA, ESO are facing similar (or higher) workload
  - Looking forward, we expect high proposal pressure for JWST
  - And TMT should likely anticipate similar pressures
- HST's system is borrowed from other observatories and missions
  - Neither perfect nor unique
  - There are some pointers for TMT

Overview

Telescope time allocation for a multi-purpose, multi-national observatory becomes

- ➔ Program scale
- ➔ Science topic
- ➔ Constituent parties

Realistic goals:

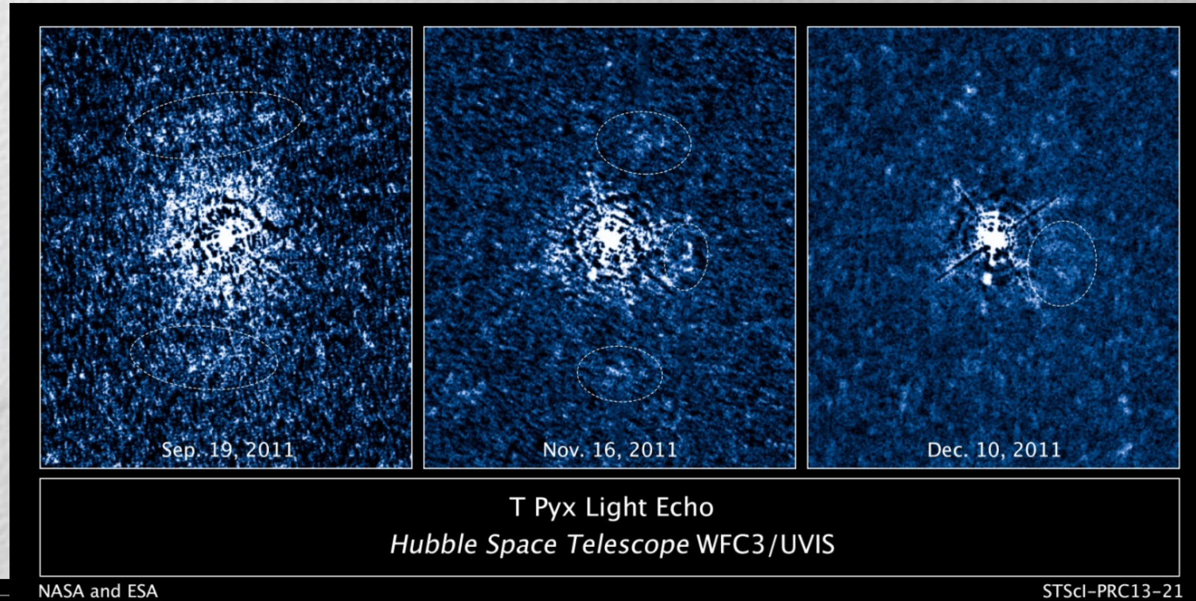
Optimise the science;
Achieve partial satisfaction for most of the constituents.

Program scale and focus

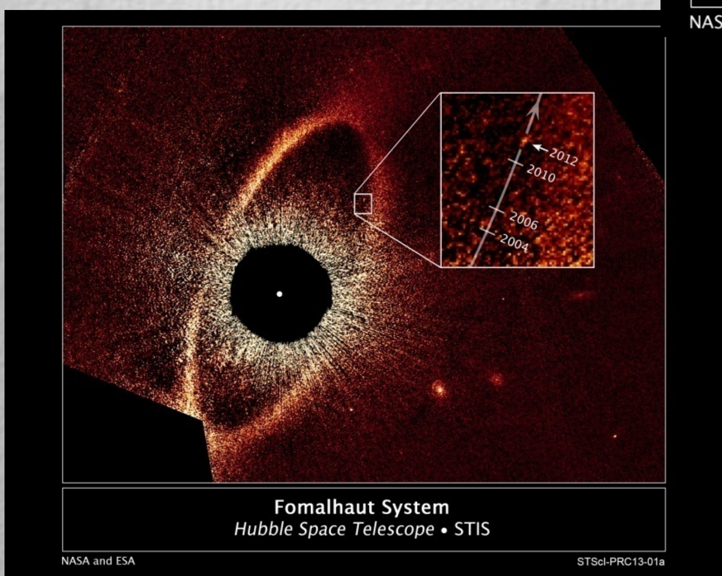
- In recent cycles,
  - ~50% of the orbits are allocated to many small programs (1-20 orbits)
  - ~30% of the orbits are allocated to a few large programs (>75-100 orbits)
  - Small programs (<20 orbits) support ~1100 investigators, with 4-5 members per team
  - Large programs(>75) support ~150-200 investigators, with 20-50 members per team
- Science
  - Small programs are generally narrowly focused experiments, targeting no more than a handful of objects to address specific questions, e.g.

Small Programs

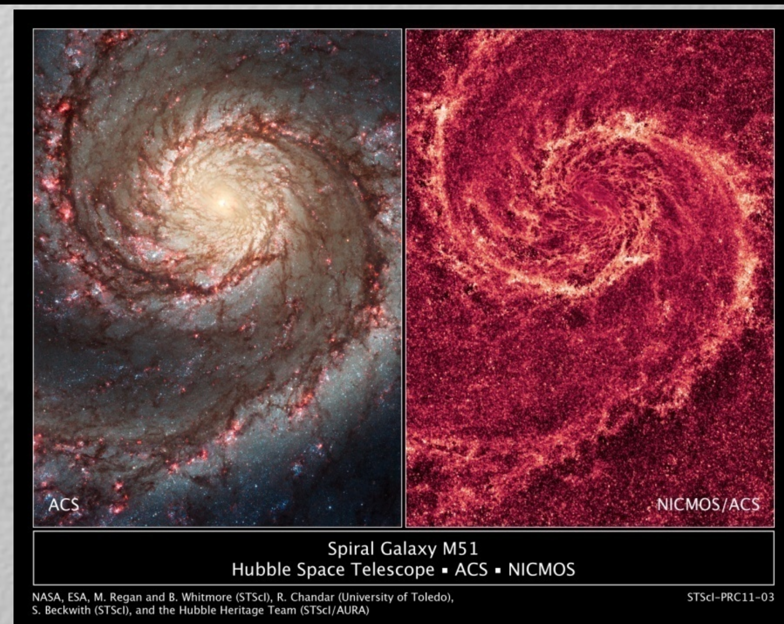
Using light echoes to map ejecta around a recurrent nova



Tracking an exoplanet orbit



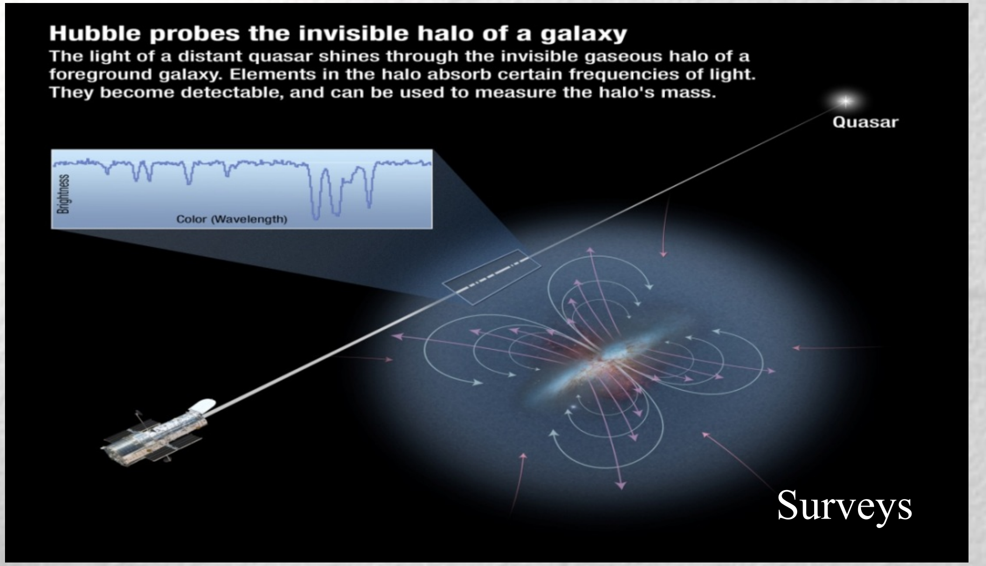
Mapping star formation in M51

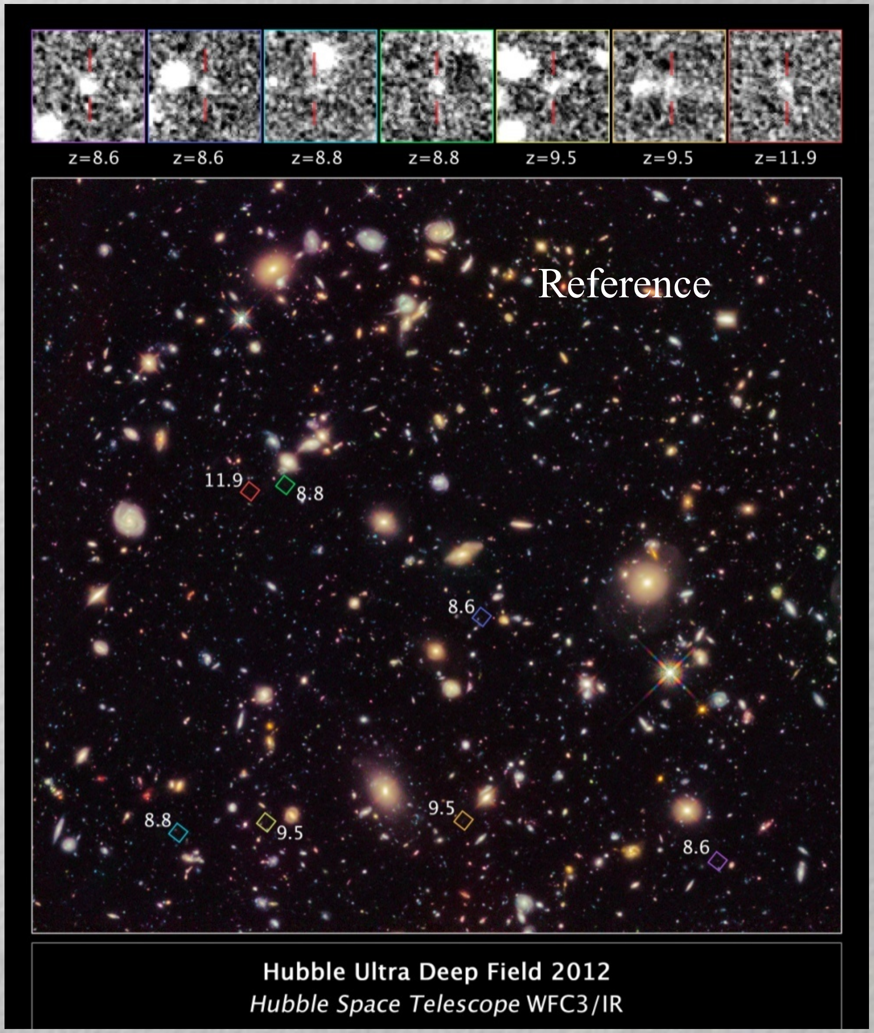


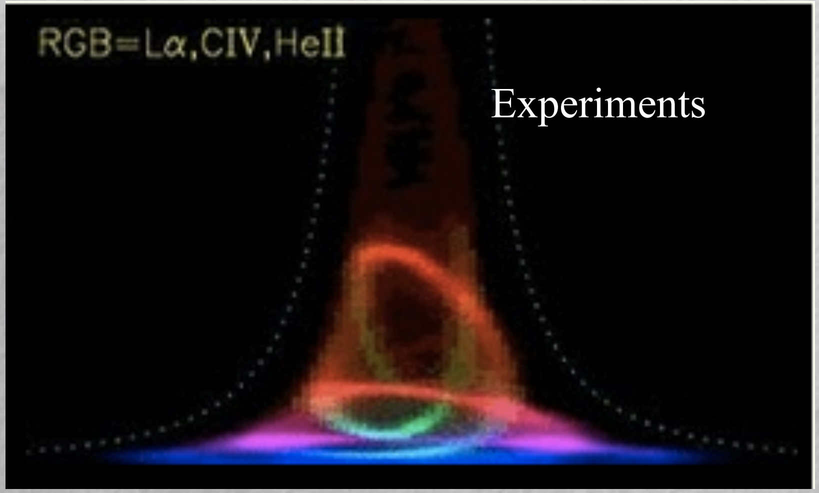
Program scale and focus

- In a typical cycle,
  - ~50% of the orbits are allocated to many small programs (1-20 orbits)
  - ~30% of the orbits are allocated to a few large programs (>75-100 orbits)
  - Small programs (<20 orbits) support ~1100 investigators, with 4-5 members per team
  - Large programs(>75) support ~150-200 investigators, with 20-50 members per team
- Science
  - Small programs are generally narrowly focused experiments, targeting no more than a handful of objects to address specific questions
  - Large programs are often surveys, covering sufficient individual targets to allow reliable statistical analysis of intrinsic properties
  - Large/Treasury programs can provide reference datasets for multiple scientific investigations, and
  - Large programs can be narrowly focused experiments that focus on large-scale issues that require extensive datasets

Large programs:







Measuring black hole masses through reverberation mapping in AGN

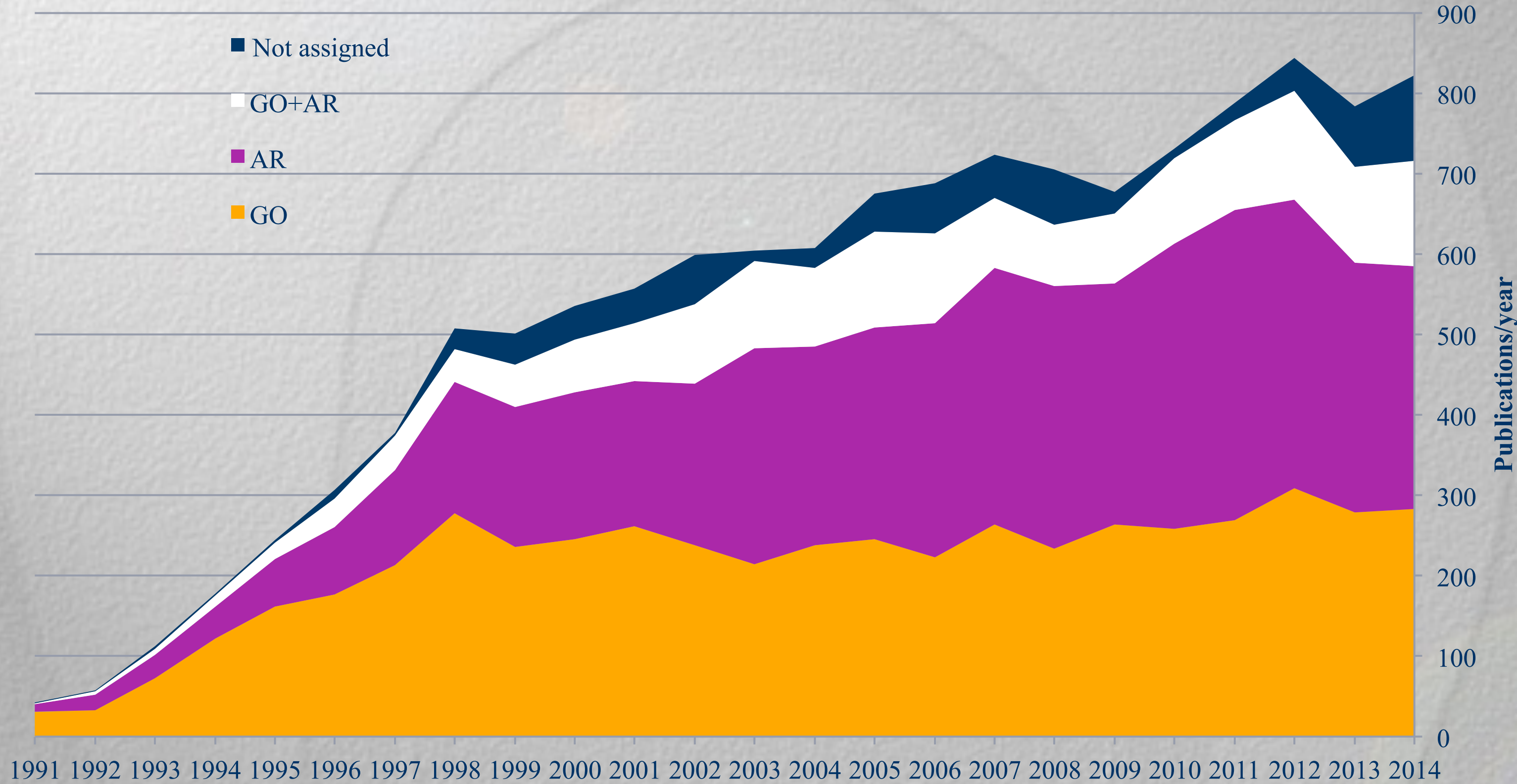
Community impact

- Analysis of publications from HST programs (Apai et al, 2010, PASP) shows that
  - Small programs produce more papers (& more citations) per orbit, but individual papers have relatively low impact
  - Large programs produce fewer papers/orbit, but more papers per program, and generally have a higher impact (more citations/paper)
  - Treasury programs generally produce more publications that Large programs
    - Archival data access is crucial in maximising the science from Large & Treasury programs

HST Publications

12,939 papers 5,249 GO papers 5,076 archival papers 1,785 GO + AR

Archival access doubles the science output



Productivity

Program	Type	Cycle	Science focus	Orbits	Publications	Notes
HDF	DD	5	Galaxy evolution	150	220	Imaging-multimission
PANS	Large	11	High-z supernovae	420	35	Imaging
GOODS	Treasury	11	Galaxy evolution	398	596	Imaging -multimission
UDF	DD	12	Galaxy evolution	400	150	Imaging-multimission
COSMOS	Treasury	13	Galaxy evolution	590	254	Imaging-multimission
PEARS	Treasury	14	Galaxy evolution	200	33	Grism spectra
UV UDF	Large	14	Galaxy evolution	204	21	Imaging-multimission
Dec_dust	Large	14	High-z supernovae	219	36	Imaging
ANGST	Treasury	15	Stellar populations	218	54	Imaging
SHOES	Large	15	High-z supernovae	208	8	Imaging
WFC3 ERS	DD	17	Star formation	214	209	Imaging, grism spectra
UDF09	Treasury	17	Galaxy evolution	193	101	Imaging-multimission
3D-HST	Treasury	18	Galaxy evolution	248	66	Grism spectra
PHAT	MCT	18-20	Stellar populations	834	39	Imaging
CANDELS	MCT	18-20	Galaxy evolution, SNe	902	207	Imaging-multimission
CLASH	MCT	18-20	Galaxy clusters, SNe	524	62 + 19	Imaging-multimission
Frontier Fields	DD	20-21 (22)	Galaxy clusters, galaxy evolution	560-840	38	Imaging-multimission

Enabling large programs

Large programs need their own space:

- Cycle 1-8: Single-phase time allocation
  - 6-15 panels ranked *all* proposals (regardless of size) within defined specialist areas (eg AGN, Binary stars, Cool Stars,…)
    - Panels allocated ~80% of the time
  - TAC (panel chairs + TAC chair) served as a merging TAC
    - Allocated ~20% of the time

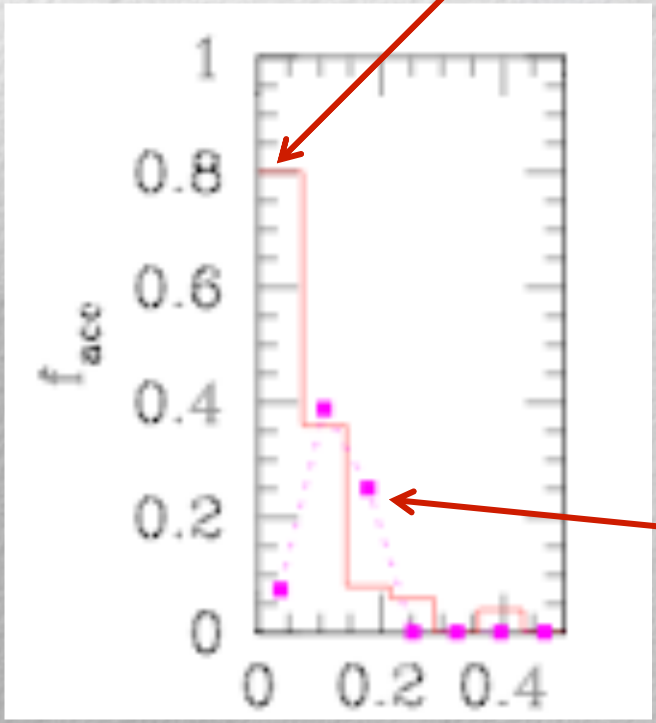
- From Cycle 1 through Cycle 8, only 5 Large (>100 orbit) programs were selected, for a total of 990 orbits (4.4% of GO allocation)
  - This includes the two Key projects

- In an unconstrained environment, peer review committees tend to favour programs that use less resources

➔ Since Cycle 9, Large/Treasury programs have a separate time allocation

Game theory and proposal size

Submitted proposals
Arbitrary vertical scaling



S_P

Each panel (including the TAC) has a fixed orbit allocation

Scale the orbits requested by each proposal against the panel orbit allocation, A_O

Proposal size, $S_P = N_{orb} / A_O$

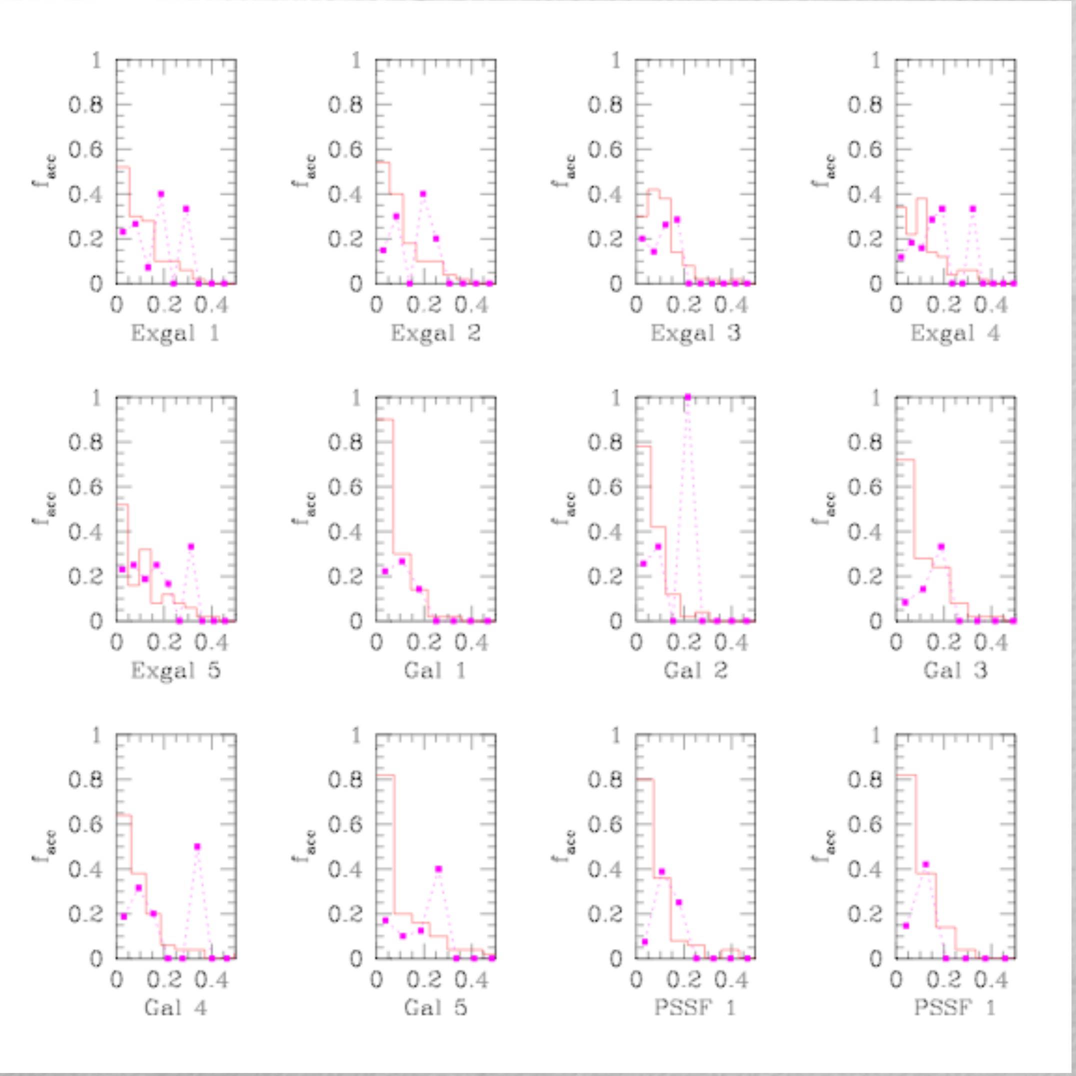
Look at the proposal success rate,

$f_{acc} = N_{accepted} / N_{submitted}$

as a function of S_P

Success rate

Constraints: Game theory and proposal size



Cycle 17 data

- Analysis of many panels shows consistent statistical behaviour
- **The success rate of a proposal drops significantly when**
- **$S_P > \sim 0.25\ A_O$**

•TAC

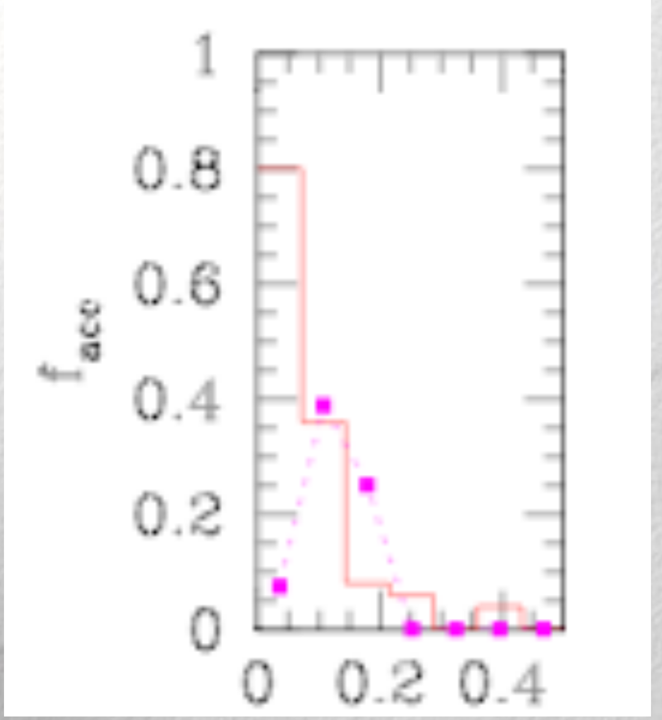
•A_O ~1000 orbits

•**S_P (max) ~ 250 orbits**

The community intuitively understand game theory; *PI's tailor their proposals to meet practical limits.*

MCT programs were introduced as a separate category to avoid this selection bias.

Game theory and panel proposals



S_P

- Each panel has a fixed orbit allocation
- Adopting **S_P (max) ~ 0.25 A_O**

  then for A_O ~100-160 orbits

•**S_P (max) ~ 35-40 orbits**

Cycles 11-20 – aimed to mitigate through an orbit subsidy from a central pool

But increasing proposal pressure demanded more panels to limit panelist workload

9 panels in Cycle 11

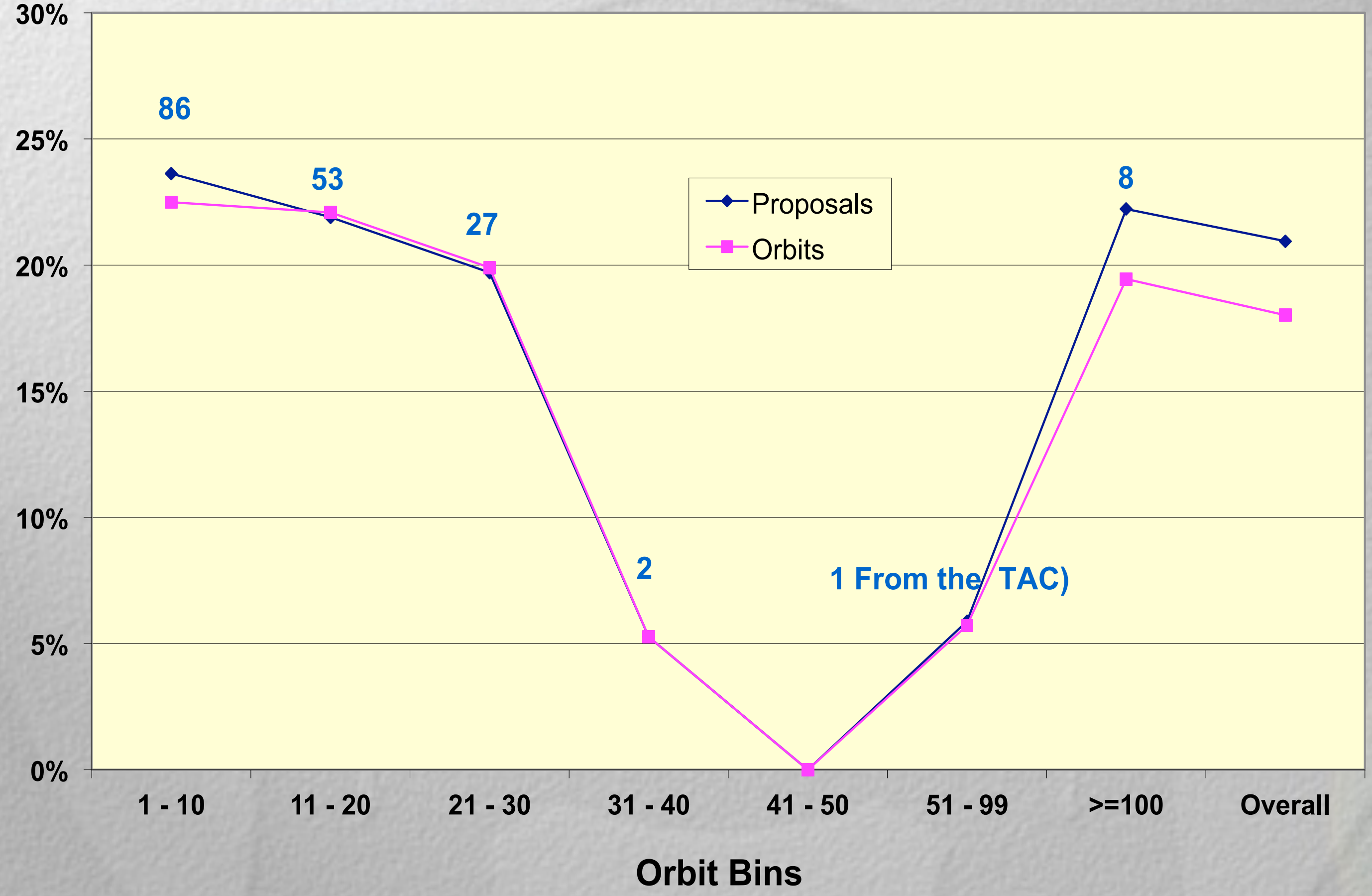
14 panels in Cycle 20

More panels, same resources → fewer orbits/panel

Into the valley of death..

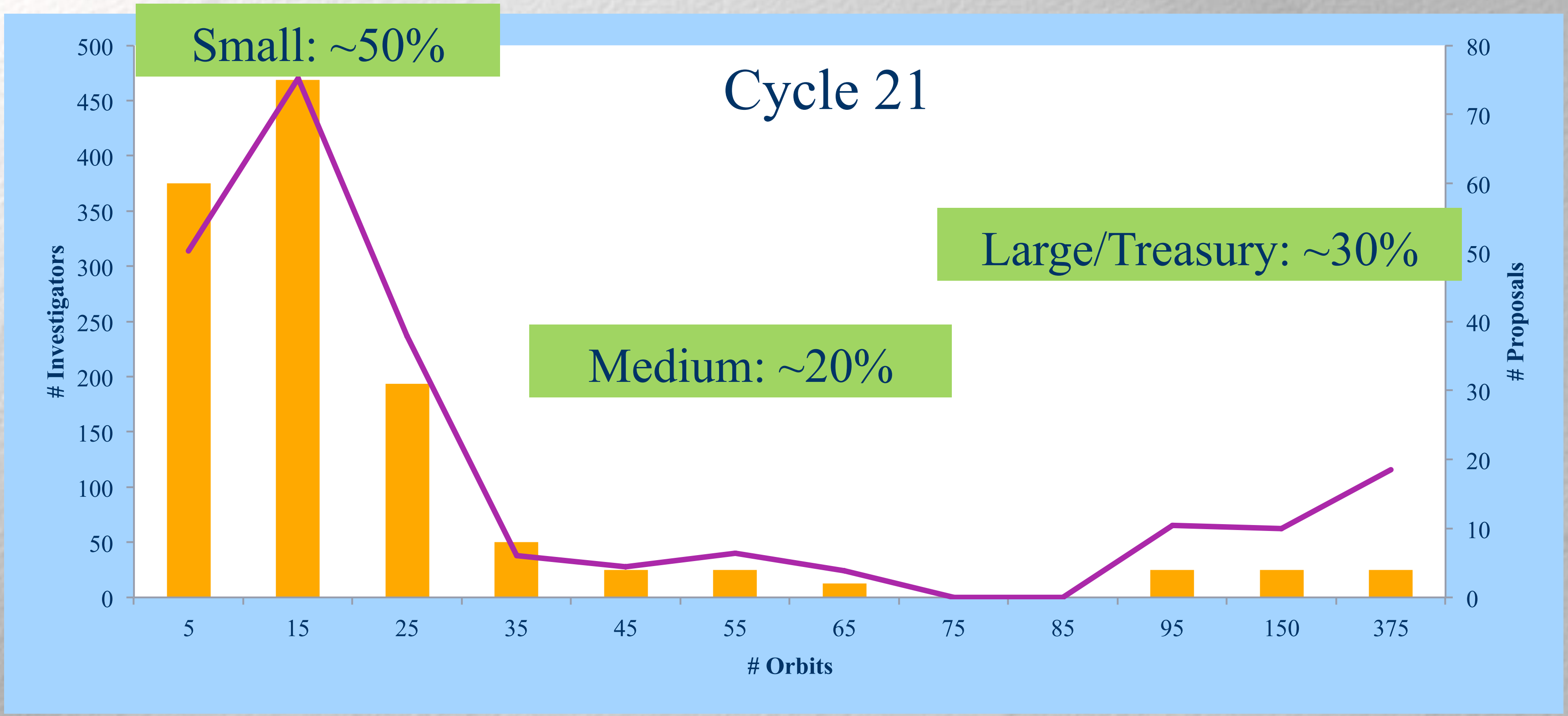
We added a medium-size category in Cycle 21:

35-74 orbits - separate orbit allocation



Cycle 20

Program size & community involvement



- Majority of programs are relatively small scale, 3-5 investigators/ program
- Small number of large programs with 20+ investigators

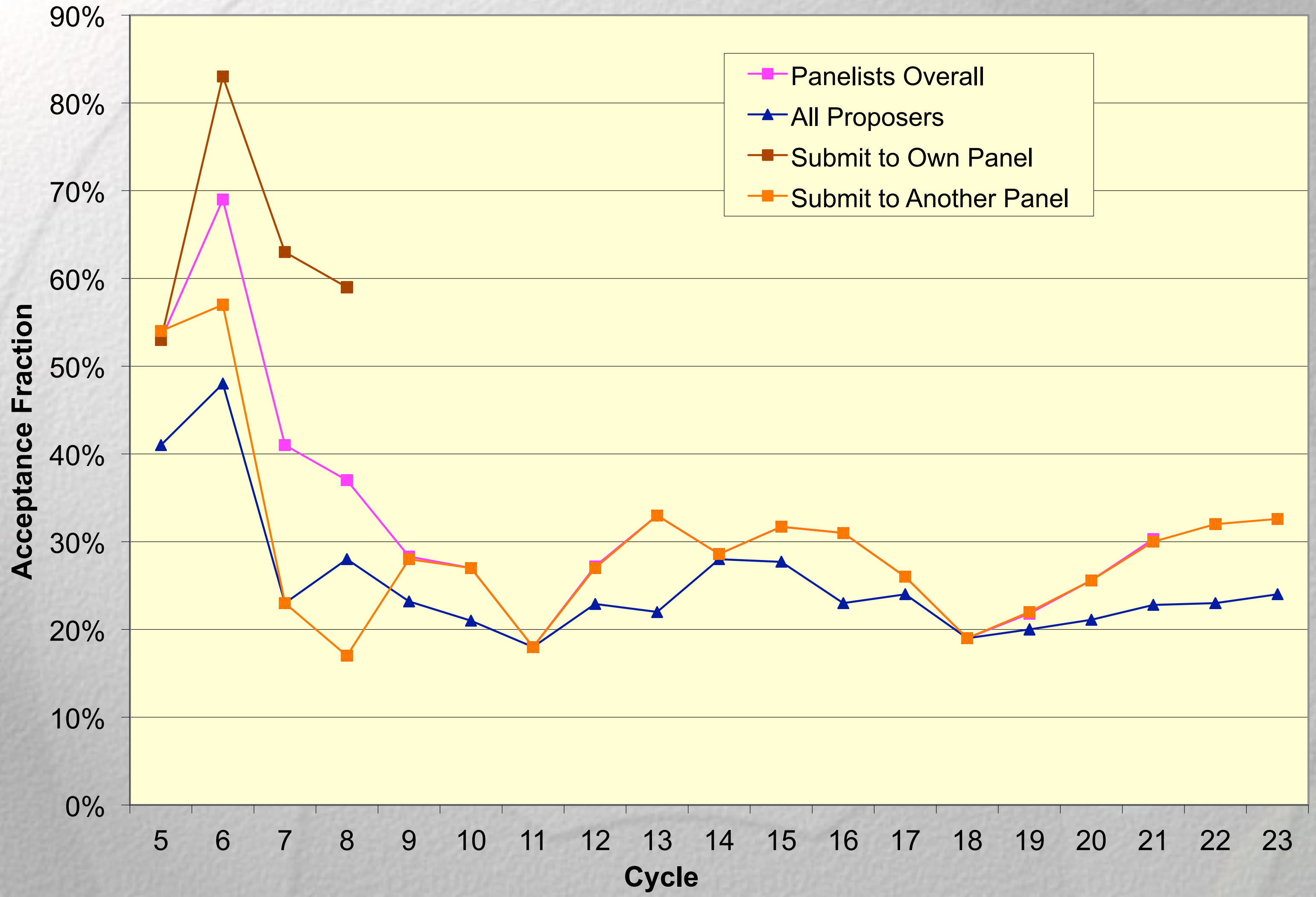
Program scale

- Observing programs of different scale tackle different types of science questions
- Balancing the size distribution generally requires external constraints
  - Committees generally like to please as many supplicants as possible
  - The TAC structure needs to take that into account

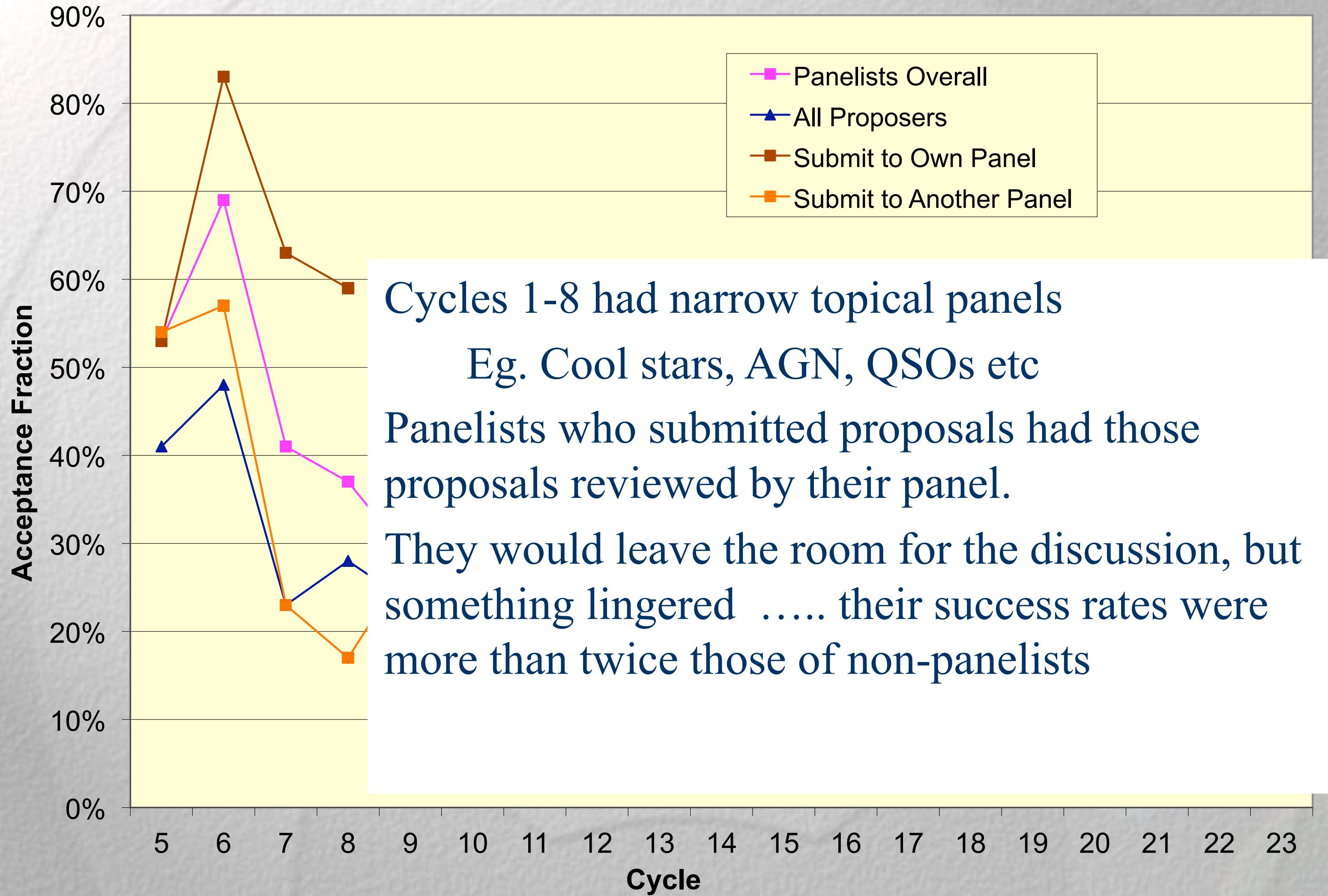
Science Topics

- How do achieve the appropriate balance for different science atreas?
- With HST, proposers self-select science categories
  - Those generally determine the panels that will review the proposals
- The current system has sets of mirror panels that cover broad topics
  - Stars, Galaxies, AGN & IGM, Planets, Cosmology, Stellar Populations

Panelist Acceptance Fraction



Panelist Acceptance Fraction



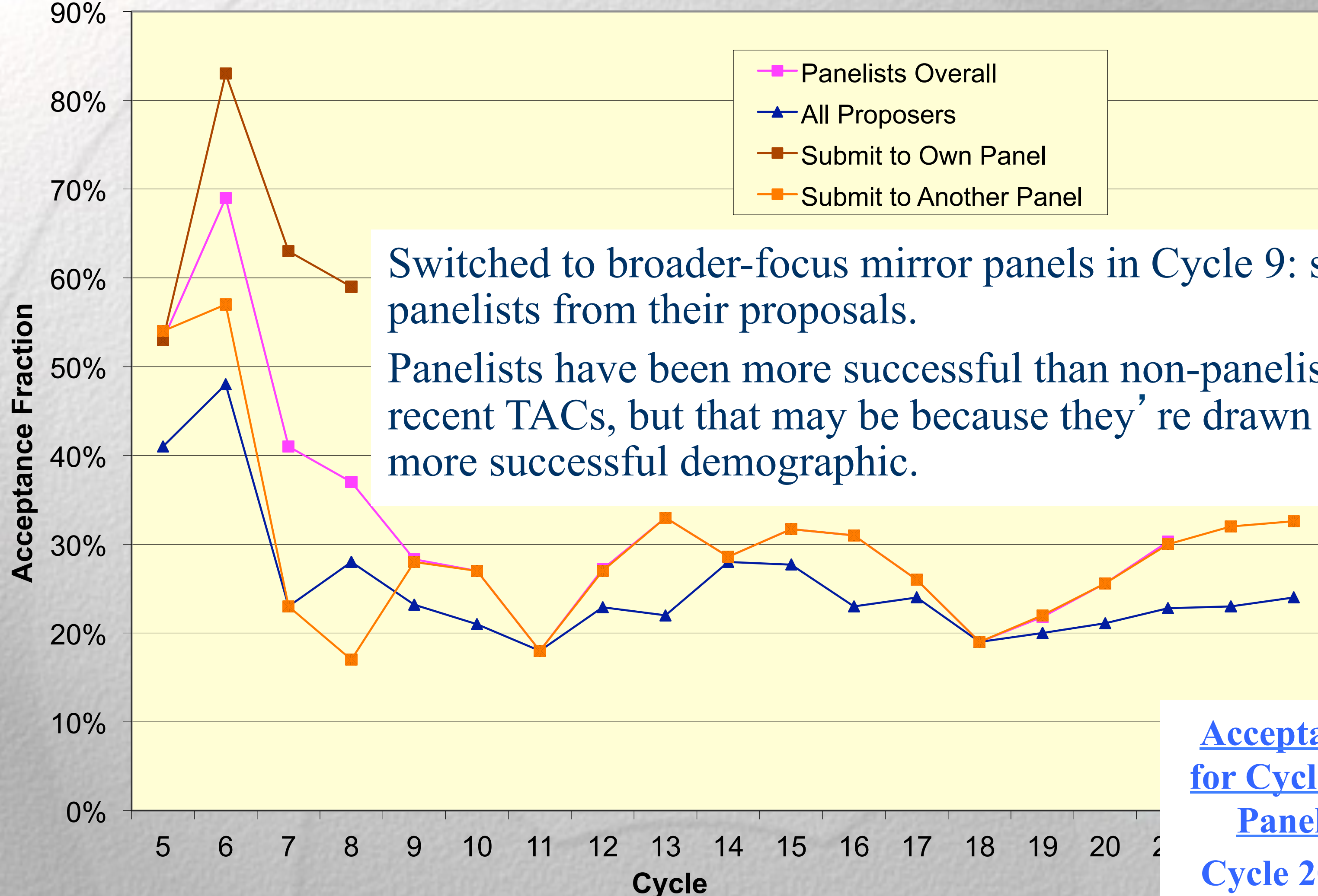
Cycles 1-8 had narrow topical panels

Eg. Cool stars, AGN, QSOs etc

Panelists who submitted proposals had those proposals reviewed by their panel.

They would leave the room for the discussion, but something lingered ….. their success rates were more than twice those of non-panelists

Panelist Acceptance Fraction



Switched to broader-focus mirror panels in Cycle 9: separate panelists from their proposals.

Panelists have been more successful than non-panelists in recent TACs, but that may be because they're drawn from a more successful demographic.

Acceptance rate for Cycle 23 TAC Panelists in

Cycle 20 is 29%

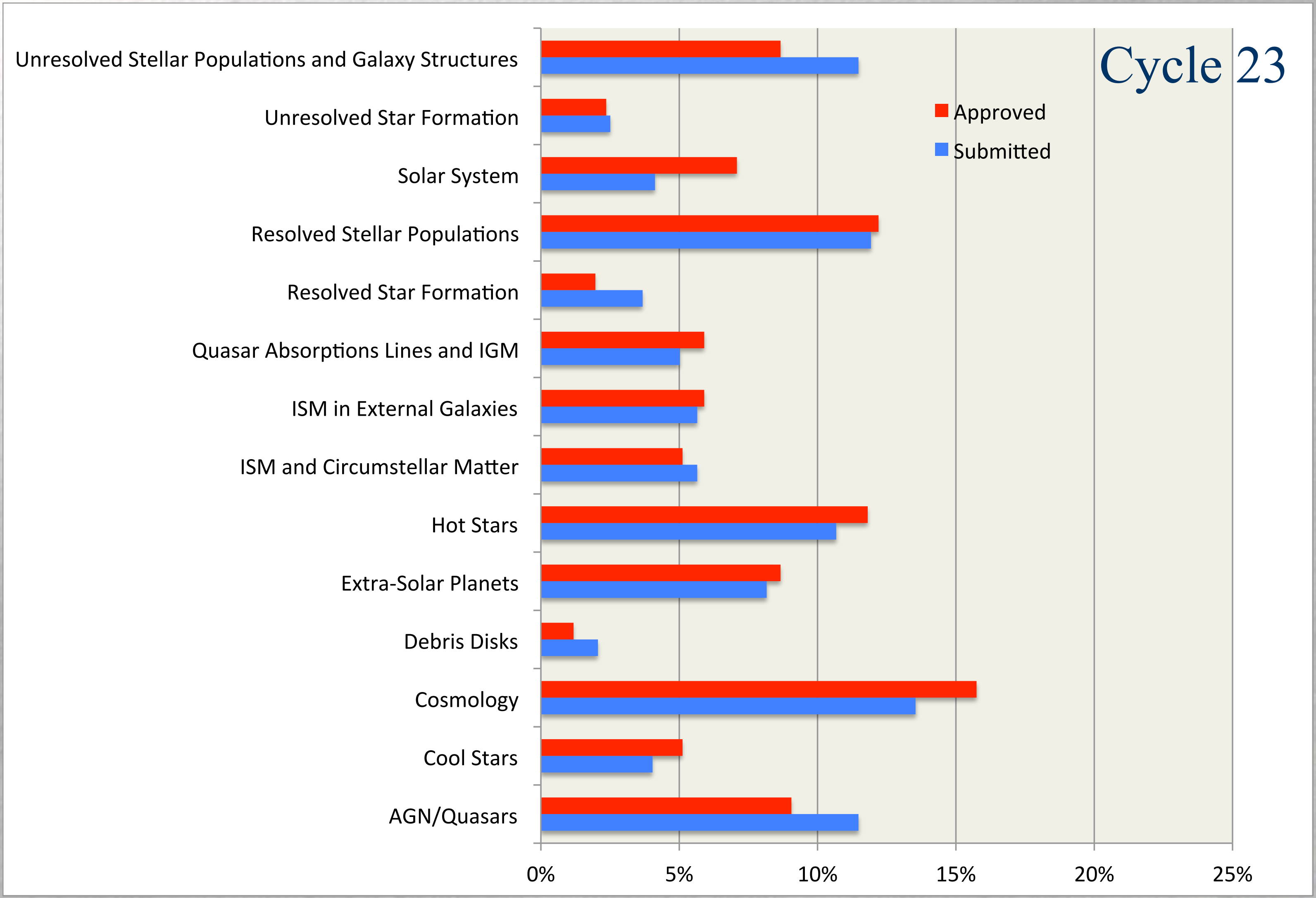
Cycle 21 is 35%

Cycle 22 is 37%

Science Topics

- How do we achieve the appropriate balance for different science atreas?
- With HST, proposers self-select science categories
  - Those generally determine the panels that will review the proposals
- The current system has sets of mirror panels that cover broad topics
  - Stars, Galaxies, AGN & IGM, Planets, Cosmology, Stellar Populations
- Panel orbit allocation are based on proposal pressure and orbit requests (equal weights)
  - Community interest tends to drive the allocations in difference science areas, with some latitude for re-balancing within a panel

Science Category Distribution for Proposals



Science Topics

- How do we achieve the appropriate balance for different science areas?
- With HST, proposers self-select science categories
  - Those generally determine the panels that will review the proposals
- The current system has sets of mirror panels that cover broad topics
  - Stars, Galaxies, AGN & IGM, Planets, Cosmology, Stellar Populations
- Panel orbit allocation are based on proposal pressure and orbit requests (equal weights)
  - Community interest tends to drive the allocations in difference science areas, with some latitude for re-balancing within a panel
- We have generally avoided emphasising specific research topics in proposal calls
  - Large DD programs have seeded some research areas for community research (Shoemaker-Levy 9, OPAL; HDF, UDF, Frontier Fields)
  - Broad initiatives for types of science: e.g. UV initiative is designed to stimulate more proposals, not support specific science topics

Picking the science

We generally tell a TAC to "pick the best science". What do we actually mean by that?

- Philip Kitcher has developed the concept of "well-ordered science" (Science, Truth & Democracy, 2001)
- "..there is no absolute standard of the significance (practical or epistemic) of research projects, nor any standard of the good apart from subjective preferences. The only non-arbitrary way to defend judgments concerning research agendas in the absence of absolute standards is through democratic means of establishing collective preferences."
  - The Social Dimensions of Scientific Knowledge, Stanford Encyclopedia of Philosophy
- That is, in this model, the fairest means of assigning priorities and resources to scientific projects is by polling the scientific community.
- Time allocation relies on advice from a subset of that community – the TAC. It is therefore important to ensure that the TAC represents a fair sampling of the whole:
  - This argues that we should aim for multiple proposal opportunities and draw TAC members from as broad a cross-section of the community as possible, i.e. maximise intellectual cycles and minimise repetition of TAC members.

Proposal Selection

- Conventional wisdom is that it's easy to select the "best" and "worst" proposals – that's not necessarily borne out by experience
  - HST panel members tend to show dispersions of ~25-30% in preliminary rankings
    - There is not unanimity on the highest and lowest ranked proposals
  - That does **not** mean that the process is random – it does mean that there is a substantial subjective component
    - **The longer a panelist serves on a TAC, the longer those subjective preferences are present**
- Rankings tend to converge following the panel discussion
  - But the dispersion is still ~15-20% in ranking
- Crucially, the exact ranking for a given proposal is not as important as whether a proposal is "accepted" or "rejected"
  - Examining HST TAC data, each panelist typically "endorses" ~2/3rds of the accepted proposals in a given panel
  - No-one gets everything, but everyone gets more than half of what they'd like

Constituents

- HST accepts proposals from the worldwide astronomical community
- HST is a collaborative venture between NASA and ESA; the ESA MOU requires that ESA scientists should be allocated to at least 15% of the observing time over HST's lifetime
  - This requirement has been satisfied naturally by the standard TAC process
- Multi-national/multi-institutional collaborations with pre-specified observation fractions are clearly more complex
  - ESO, ALMA & Gemini have direct experience in these areas and can offer more cogent advice
- A single time allocation process is likely to be most effective for planning large-scale, co-ordinated programs

Summary

- Programs on different scales tackle different types of science
  – A balanced program distribution requires an appropriate TAC structure
- Let community interest drive the topical choice within the science program
  – Draw TAC members from as broad a cross-section of the community as possible
  – Don't "lock in" individual TAC members for long terms of service
- As a multi-national/multi-institutional observatory, TMT may need to organise multiple TACs
  – More coordination will likely help maximise the science from larger-scale programs

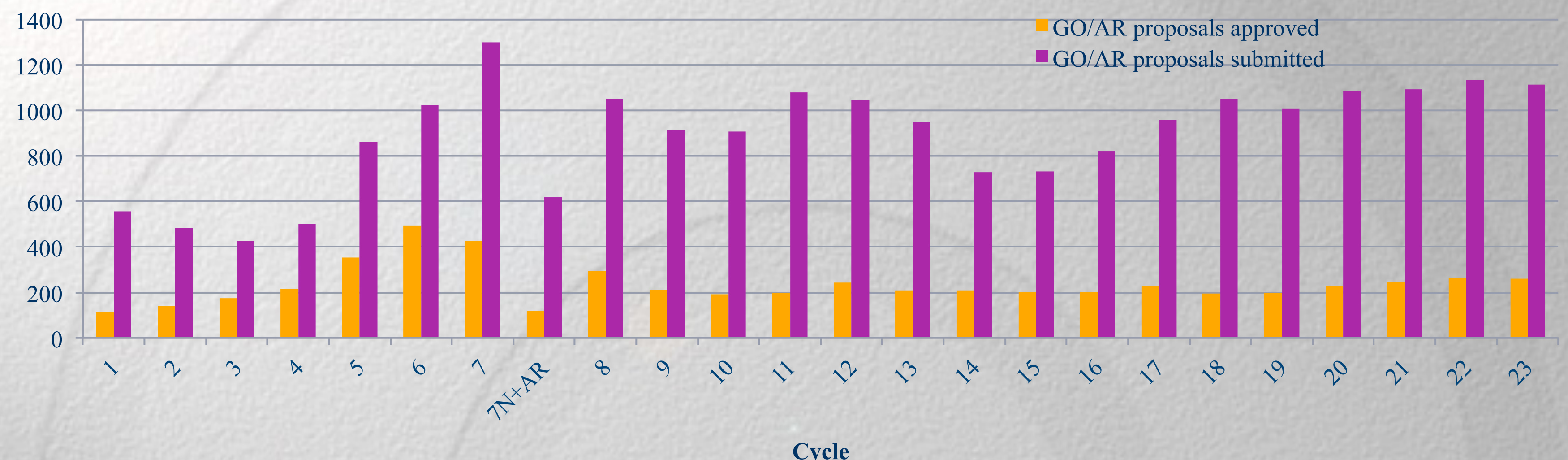
Maybe ~5-7 Years from now

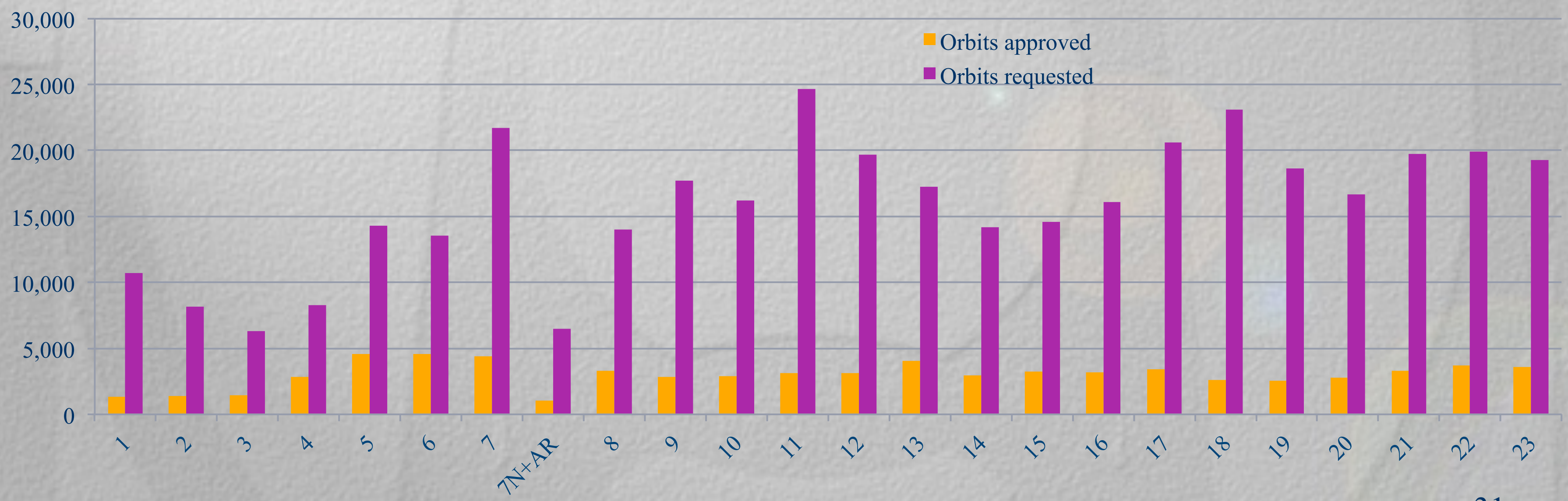
HST →

← TMT

Backup

Proposal pressure



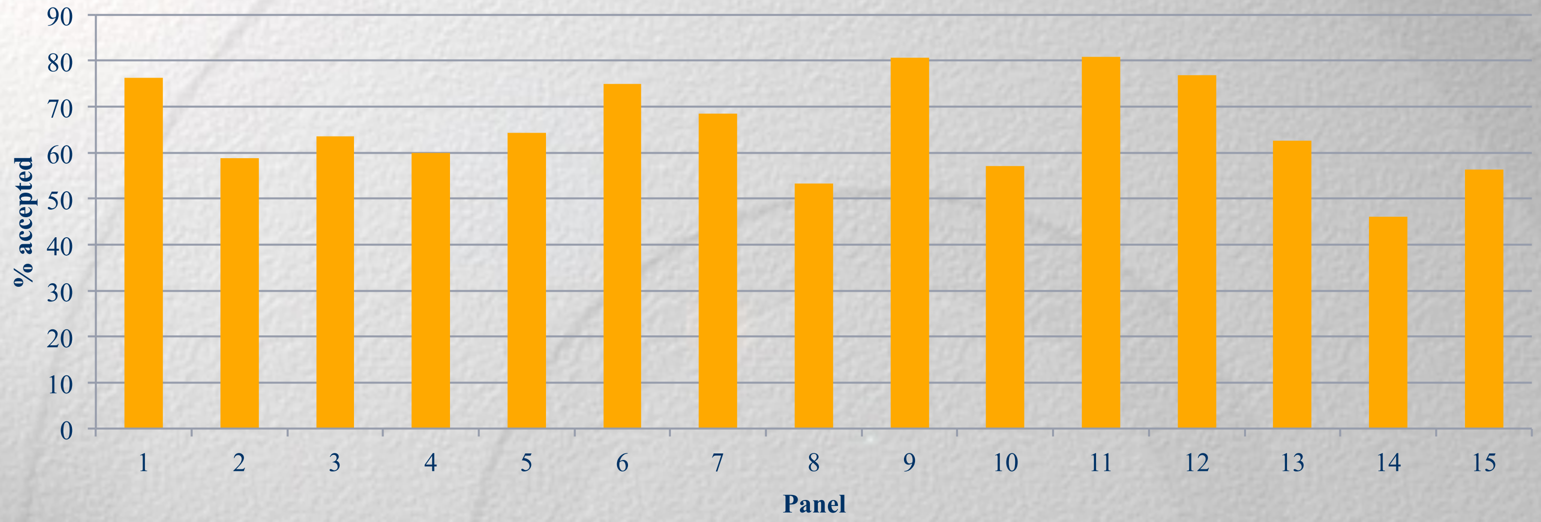


How well do individual panelists agree after the discussion?

As with the preliminary grades, we can compare the final ranked list against the results from individual reviewers

Overall, the agreement is closer, but significant differences remain in the rankings by individual reviewers.

How well do the preliminary and final ranked lists agree?



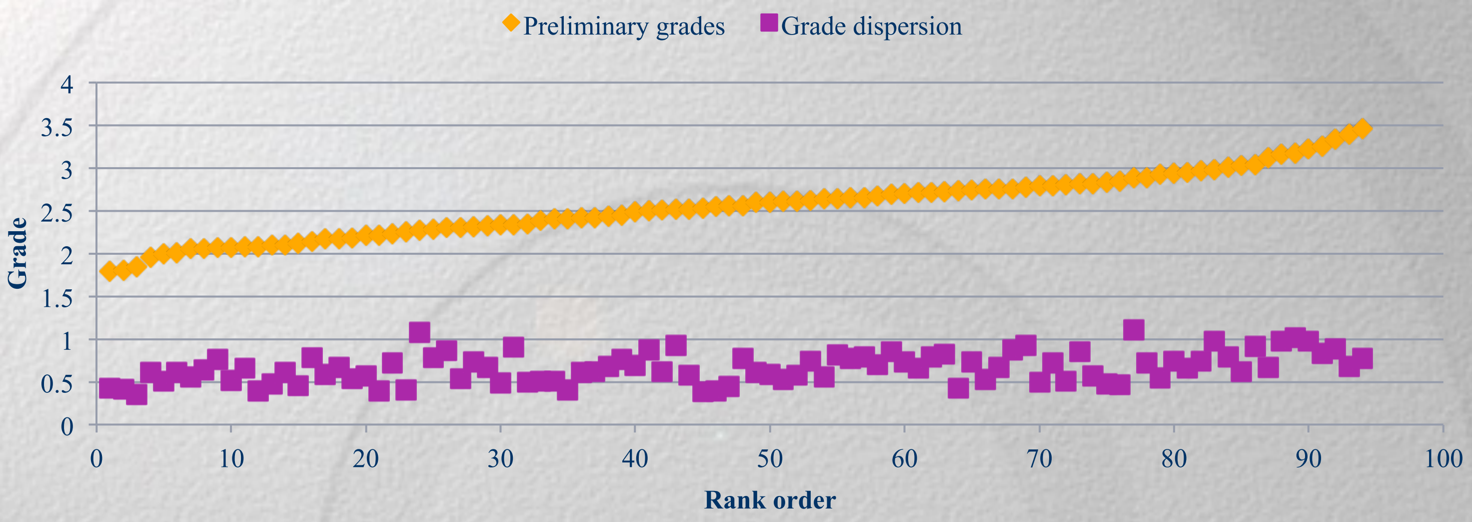
Each panel allocates time to N proposals

- What fractions of those proposals would have been awarded time had we used the preliminary grades to select accepted proposals?

Overall, 252 proposals were accepted in Cycle 21; 170 (67%) would have been accepted based on the preliminary ranking

- The overlap ranges between ~45% and ~80% for the individual panels

Dispersion in average grades

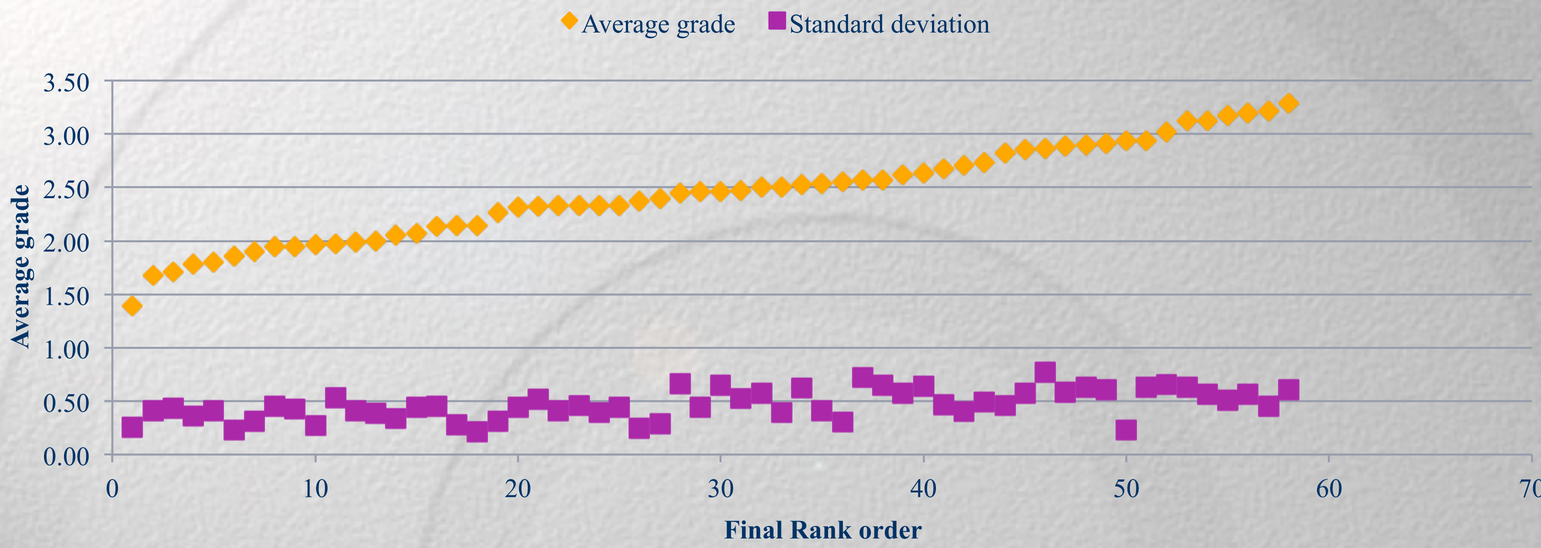


Panelists are asked to provide preliminary ranks from 1-5 for proposals, where 1=good, 5=poor.

We do not impose a particular system, but ask that panelists use the full range available.

The dispersion in grades tends to be lower for highly-ranked proposals, and increases (slowly) towards lower rankings; there is significant dispersion

Dispersion in final grades



Dispersions for the proposals ranked by Panel X

Note that ~35 proposals were triaged

Overall, the dispersions decrease showing greater agreement among the panelists, with a milder trend to increased dispersion at lower ranks.

However, only a handful of proposals have $\sigma < 0.3$