AIDS TO THE TAC PANELIST SELECTION

The Panel Auto-Catagorizer and Manager (PACMan)

STUC, November 2015, L. Strolger

REVIEW OF "CATEGORIZATION" AND THE PANELS

▸ Proposals are distributed to ~14 panels, organized in ~6 proposal categories (each w/ one or two mirrors)

▸ User-selected Science/Proposal categories and keywords guide proposal distribution

▸ Panelists (and Chairs) are chosen ahead of time to fill expertise in these ~6 proposal categories

PANELIST SELECTION

▸ The members of the SPG are assigned panels outside their area of expertise to reduce conflict of interests

▸ Rely on limited base of information to fill panels: past STScI service, colleague recommendations, group discussions, etc

▸ Time consuming; potential biases in using/reusing a select set of HST users

PANELIST SELECTION

▸ Rely on panelist-selected science keywords to define areas of expertise within proposal categories

▸ Keywords can be vague, misunderstood, or misused
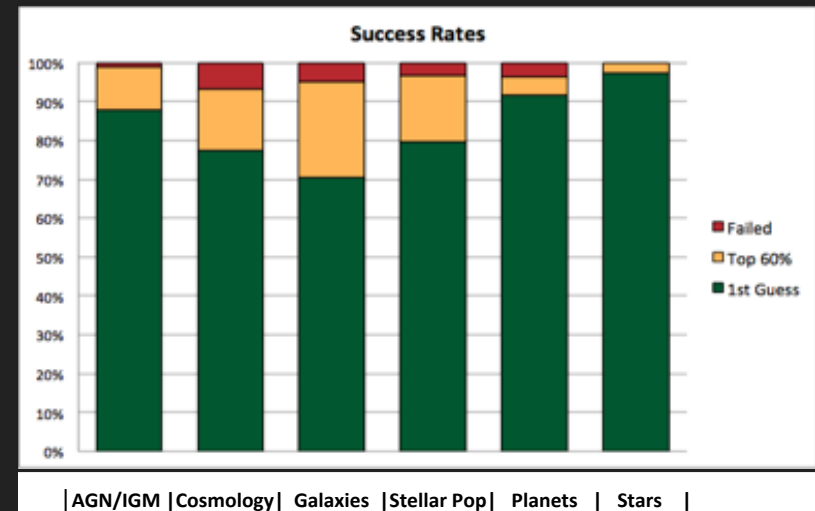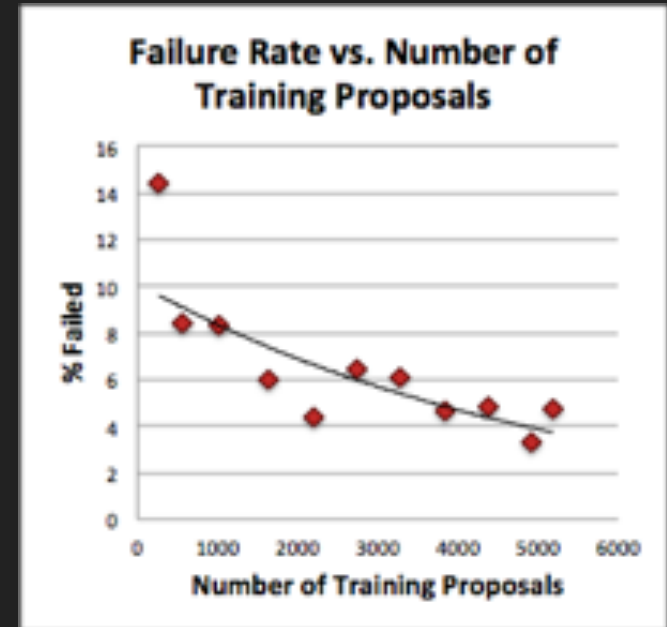
TOWARDS AN AUTOMATED PANELIST SELECTION

▸ Panel Auto Categorizer and Manager (PACMan)-- Sophia Porter (SASP 2016, JHU), Lou Strolger, Jill Lagerstrom, and Sarah Weissman (STScI).

▸ Based on tool used to categorize astronomers for the state of the profession analysis, NWNH 2010

▸ Naive Bayesian routine, where tokens are words in abstract bibliography (over last 10 years via ADS) for potential panelists

▸ Tokens are sorted into the six pools, or panel categories

TOWARDS AN AUTOMATED PANELIST SELECTION

▸ 85-95% accuracy, based on Cycle 23 panel selection (some catastrophic failures at the ADS interface)

▸ <u>Eliminates need for science keywords</u>

▸ Needs Master list (e.g., mailing list); PACMan returns lists of people best-suited for a given panel, and list alternatives satisfactory for said panel

▸ Using now to guide our initial panelist selections from <u>wide list of >8000 people</u>

PROPOSAL CATEGORIZATION

- Science categories and keywords guide a proper review. But rely on proposers to categorize their own submissions.

- Keywords can be unclear, misunderstood, or misused, and categories evolve as new fields emerge or wane. Takes time for SPG to review, correct.

- PACMan categorizes to >95% precision

Failure Rate vs. Number of Training Proposals

Success Rates

|AGN/IGM |Cosmology| Galaxies |Stellar Pop| Planets | Stars |

STUC, November 2015, L. Strolger

FUTURE APPLICATIONS

▸ With a better algorithms for joint probabilities (inter-dependence of words), and a larger corpus,

- we can construct more generalized "panels" of common-themed proposals

- may match/select panelists for a given submission