

# Proposal Review Rankings: The influence of reviewer discussions on proposal selection

Lisa J. Storrie-Lombardi\*, Nancy A. Silbermann, Luisa M. Rebull, Seppo Laine, Megan Crane  
Spitzer Science Center, California Institute of Technology, MS 314-6, 1200 E. California Blvd.,  
Pasadena, CA 91125 USA

## ABSTRACT

The telescope time allocation process for NASA's Great Observatories involves a substantial commitment of time and expertise by the astronomical community. The annual review meetings typically have 100 external participants. Each reviewer spends 3-6 days at the meeting in addition to one-two weeks of preparation time, reading and grading proposals. The reviewers grade the proposals based on their individual reading prior to the meeting and grade them again after discussion within the broad, subject-based review panels. We summarize here how the outcome of the review process for three Spitzer observing cycles would have changed if the selection had been done strictly based on the preliminary grades without having the panels meet and discuss the proposals. The changes in grading during the review meeting have a substantial impact on the final list of selected proposals. Approximately 30% of the selected proposals would not have been included if just the preliminary rankings had been used to make the selection.

**Keywords:** Spitzer Space Telescope, Proposal Peer Review, Telescope Time Allocation

## 1. INTRODUCTION

The proposal selection process for NASA's Great Observatories works in a similar fashion for the Spitzer Space Telescope, Hubble Space Telescope and Chandra X-Ray Observatory. Broad science review panels organized by science topic review the proposals. The individual reviewers grade the proposals in advance and then the panels meet in person to discuss the proposals and provide a final ranking for the programs within the panel. Care is taken to minimize any conflicts of interest and maximize the quality of the science review. The sheer number of proposals handled in these reviews (600-1000) and the breadth of the science proposed makes it impossible to perfectly match the expertise of the reviewers to the proposals. The call for proposals specifies that proposals will be read by broad, topical panels and should be written so that non-experts understand why the science is important.

The Spitzer Science Center (SSC) has responsibility for the selection of proposals for the Spitzer mission and the proposal selection and review process follows closely the processes successfully developed for selection of Hubble and Chandra proposals. More details of the full Spitzer proposal selection process are discussed in Laine et al. (2006)<sup>1</sup>. The review meetings last a week and involve substantial investments in time from the science center (~25 people) and the community (~100 reviewers) as well as financial resources for the travel of the reviewers and the meeting at a large hotel. Without including the cost of labor the Spitzer reviews cost from \$160-\$250k to support, i.e. the cost of more than one FTE. When asking referees to participate in the review process the success rate is only about 40-50%, i.e. to get 100 reviewers to serve approximately 250 have to be invited. One of the main reasons people say no is the travel and time commitment for the review meeting.

The Spitzer project has proposed to NASA to continue operating the observatory after the cryogen runs out (~April 2009) in what we call the 'Warm Mission'. In an effort to substantially reduce the operating costs of the observatory the expensive annual proposal review meetings must be replaced with a process that retains the scientific integrity and quality of the selection process without the large face-to-face review. To explore the impact of the review meeting itself on the outcome of the selected proposals we have estimated what the selected program would have been in Cycles 3-5 if we had just used the preliminary grades to provide the final ranked list of programs.

---

\* [lisa@ipac.caltech.edu](mailto:lisa@ipac.caltech.edu)

## 2. SUMMARY OF THE REVIEW PROCESS

### 2.1 Review Panels

The reviews for Cycles 3-5 were composed of 11 topical science panels and a time allocation committee (TAC) that consisted of a TAC chair and the 11 panel chairs. In Cycle-4 the TAC also included two ‘at-large’ members that did not participate in the panel review discussions. The panels cover broad scientific categories the distribution of science topics assigned to the panels is summarized in table 1.

Table 1. General scientific make-up of the Spitzer review panels. All science topics have parallel panels to help minimize conflicts of interest, except for the solar system panel.

<b>SPITZER REVIEW PANELS – SCIENCE TOPICS</b>	
<u>Extragalactic</u> Panels 1A,1B,1C Panels 2A,2B	distant universe, high-z galaxies/clusters, AGN nearby Universe, low-z galaxies/clusters, local group
<u>Galactic</u> Panels 3A,3B,3C Panels 4A,4B	ISM, Galactic structure, evolved stars, compact objects, star formation circumstellar disks, brown dwarfs, extrasolar planets
<u>Our Solar System</u> Panel 5	asteroids, small bodies, planets, zodiacal background, comets

The panels are composed of 6-10 members, including the chair. The extragalactic and galactic panels typically review 60-85 proposals and the solar system panel reviews 35-45 proposals. Panel members are recruited in advance of the proposal deadline based on the science mix submitted in the previous cycle. While we primarily try and recruit more senior members of the community we generally have at least one post-doc per panel. For the review membership we aim to balance geographical location, institution size and gender of the review members. We recruit approximately 20% of the reviewers from outside the U.S. While members of the SSC, the Infrared Processing and Analysis Center, the Spitzer Project Office and the Spitzer instrument principal investigators may not serve on the review we do recruit members of the guaranteed time observer teams. Scientists with topical expertise but no connection to Spitzer are also asked to serve. We had 88, 97 and 98 reviewers participate in the cycle 3, 4 and 5 reviews, respectively.

### 2.2 Proposal Types

Spitzer solicits general observer (GO) and Legacy-GO proposals, as well as guaranteed time observer (GTO) proposals, and archival (AR) and theoretical (TR) research proposals. These are summarized in table 2. Spitzer typically receives 600-800 proposals per observing cycle.

Table 2. Spitzer proposal categories.

<b>SPITZER PROPOSAL CATEGORIES</b>		
<u>GENERAL OBSERVER</u>		
Small < 50 hours	Medium 50-200 hours	Large > 200 hours
<u>LEGACY GENERAL OBSERVER</u>		
Medium or Large, strong science case plus ‘legacy’ value for the astronomical community		
<u>GUARANTEED TIME OBSERVER</u>		
MIPS, IRS and IRAC instrument PIs were each awarded 5% of the time in Cycles 3-5 but they did have priority for targets. Any duplication between proposed GTO and GO programs are resolved based on final rankings in the review.		
<u>ARCHIVAL AND THEORETICAL RESEARCH</u>		
Requests for data analysis funds only with no observing time requested.		

Starting with Cycle-3, the GTOs submitted proposals for their guaranteed time. For cycles 1 and 2 the GTOs had priority for targets and therefore submitted their programs prior to the issuance of the GO calls for proposals. The numbers of proposals submitted of each type in each cycle is illustrated in figure 1.

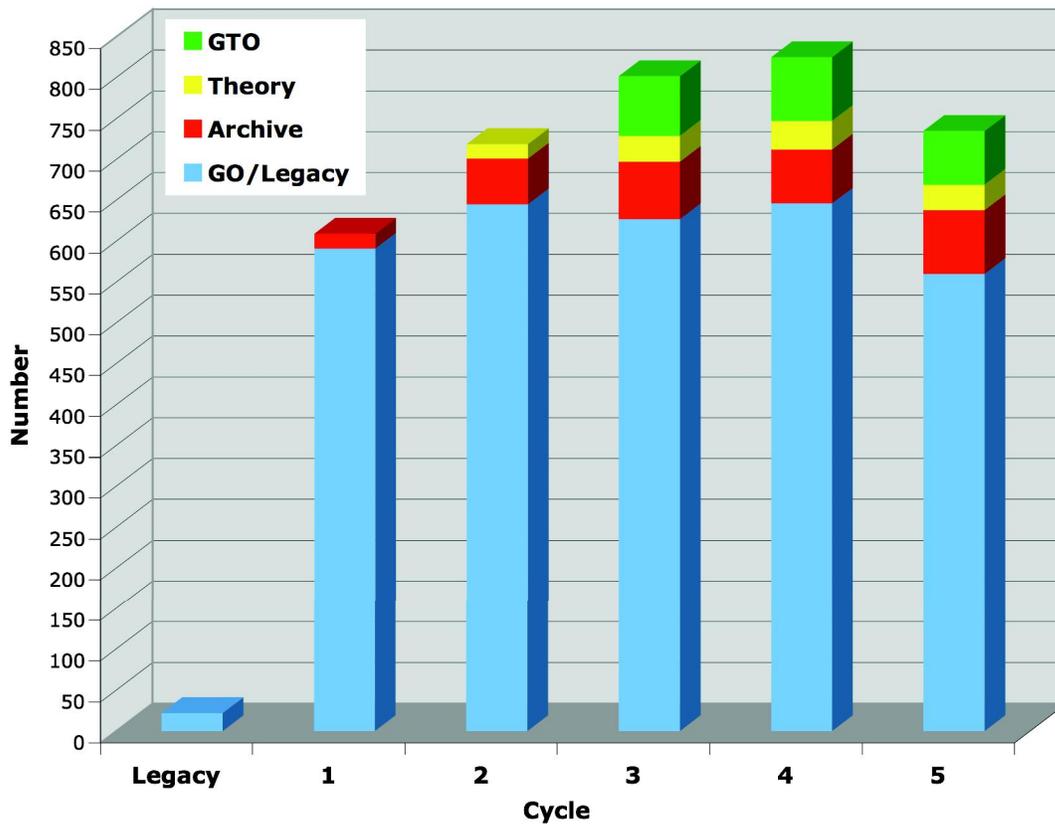


Fig. 1. Spitzer receives ~600-800 proposals per cycle. The figure shows the number of proposals in each category submitted per cycle. The original Legacy call was limited to those large programs. Cycle-1 did not solicit GO-large or Legacy programs. Cycle-5 did not solicit GO-large programs and offered half as many GO/Legacy hours as awarded in cycles 3 and 4.

### 2.3 Proposal Grading and Selection

The proposals and reviewers are assigned to panels to minimize potential conflicts of interest and maximize the match between the reviewer's expertise and the science proposed. The reviewers have approximately one month to read the proposals before the review meeting. Proposals of all types are assigned to the panels for review and it is the task of the panel to provide a final rank ordered list for all of their assigned proposals. Each proposal is assigned a primary and secondary reviewer from the panel. These people have primary responsibility for leading the discussion about the proposal at the review meeting. All reviewers are expected to read and grade all proposals in their panel (except those for which they or the SSC identify a conflict) and provide the SSC with preliminary proposal grades on the Thursday before the review meeting starts. 95-98% of the reviewers succeed in submitting their preliminary grades by this deadline. Prior to cycle-3 we did not ask the entire panel to provide grades prior to the review so we don't have preliminary grades to compare with the final rankings for cycles 1 and 2.

Prior to the review meeting the SSC triages the bottom third of the proposals in each panel based on the preliminary grades. The triaged proposals are generally not discussed at the review meeting. The review panel can 'rescue' any proposal from triage that they want to discuss. This usually happens for a few proposals that have a wide dispersion in the preliminary grades. In cycles 3-5 we have identified two proposals out of the ~2300 submitted and ~750 triaged that were rescued from triage, discussed and received final grades high enough to be selected.

The non-triaged proposals are discussed by the panels and regraded by all non-conflicted reviewers. When the panel has completed their discussions a ranked list is prepared based on the average final grades. At this stage in the process the panels can change the ranking of any proposals based on additional discussion of the science. The end result is a final ranked list of proposals. The panels are provided with an allocation of hours that directly corresponds to the GO-small proposal pressure in their panel. For example if the oversubscription factor for GO-small proposals is 5 and the panel has 1000 hours of GO-small proposals then the panel allocation will be 200 hours for GO-small programs. The small proposals in their ranked list will be selected down to the proposal that fills their hour allocation. The GO-medium, large and Legacy proposals with ranks higher than the lowest selected GO-small program will be forwarded onto the TAC for further discussion and down selection. The Archive (AR) and Theory (TR) proposals with ranks higher than the lowest selected GO-small program will be selected and forwarded onto the TAC for concurrence. The relative ranking of the GTO proposals is utilized to resolve any conflicts with selected GO programs (higher ranking program normally gets the observations) and to assign scheduling priorities.

Though the panels are instructed to just do the ranking based on the science and not to cut programs to squeeze additional ones into the selected list, once the preliminary ranked list is done and the panel can see where the selection line is drawn some proposals are moved in the rankings. For example, if an AR program that the panel really likes is just below the selection line its rank may be raised so that it falls into the selected list. If the final GO-small program straddling the selection line will put the panel more than 2% over their allocation and there is a slightly smaller program, or two small programs, ranked just below, the ordering of these might be reversed to select the two that will fit. Proposals might also be moved to create a better science balance for the panel. There are no quotas for scientific categories but if a panel discovers that they haven't selected any programs in a topic they wanted to support this also might lead to slightly reorganized rankings.

For the full cryogenic cycles (2,3,4) 5600-6000 hours of GO/Legacy time were available. The Legacy programs executed in cycle-1 were selected in a separate call for proposals; therefore only 3700 hours of GO time was available in cycle-1. Cycle-5 is not a full cryogenic cycle and 2800 hours of science priority 1 and 2 observations were selected. An additional 2200 hours of priority 3 observations were also selected but these are only likely to be executed if the cryogen exceeds the nominal expected lifetime. The priority 3 programs are not included in the statistics provided in the next section. The requested and awarded GO/Legacy hours, archival dollars and theory dollars for all Spitzer observing cycles are shown in figures 2 and 3.

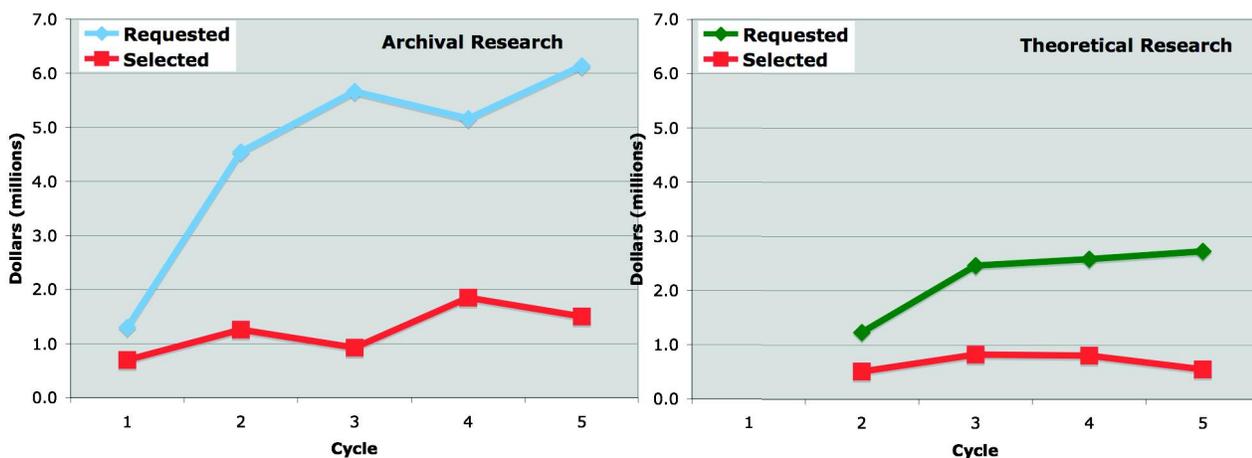


Fig. 2. The requested and selected dollars for the Archival Research Proposals (left) and Theoretical Research Proposals (right) are shown. Archival programs were introduced in Cycle-1 and theory programs were introduced in Cycle-2.

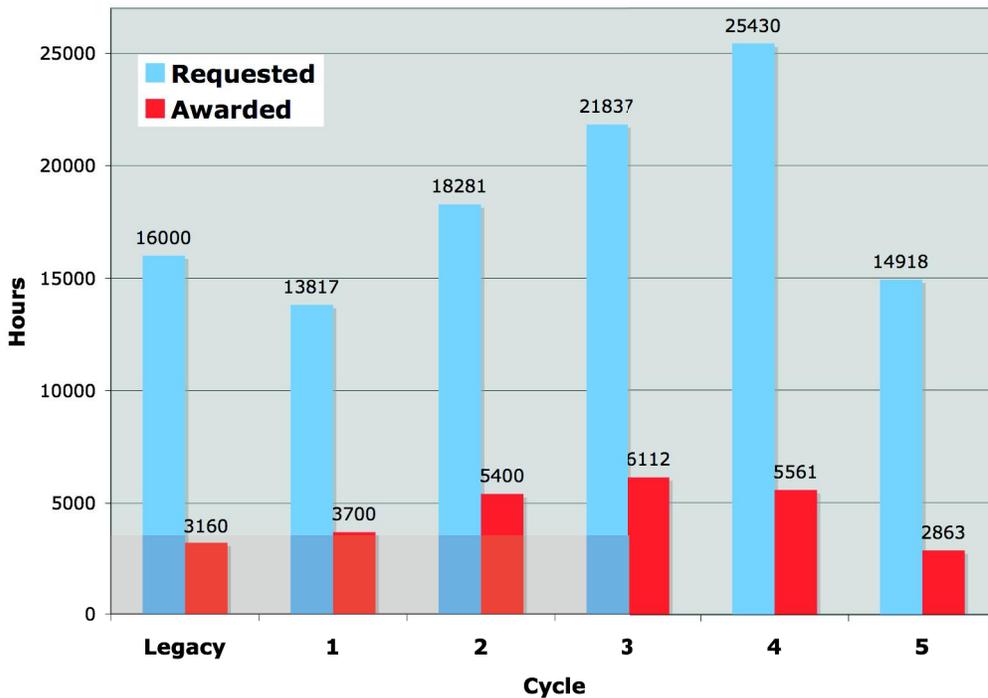


Fig. 3. The requested and selected hours for GO/Legacy programs for the original Legacy call and Cycles 1-5 are shown. The programs selected in the original Legacy call were executed during cycles 1 and 2. Cycle-5 is not a full cryogenic cycle and the selected science shown is for the priority 1 and 2 observations. An additional 2200 hours of priority 3 observations were also selected in Cycle-5 for execution if the cryogen exceeds the nominal expected lifetime.

### 3. RESULTS

To estimate the proposals that would have been selected if only the preliminary grades had been considered we have determined the rank of the lowest ranked GO-small proposal that was selected in each panel and then identified the proposal with that rank in list of preliminary average grades. For example, if the 15<sup>th</sup> ranked proposal was the selection line in the final ranking for a panel then we compared the top 15 proposals in the final list with the top 15 proposals determined just from the preliminary grades. This is not a perfect comparison because we aren't specifically confirming that the number of hours selected in the preliminary list down to a particular rank matches that in the final list. For our purposes though, determining how much the rankings changed based on the review meeting, the results should still be valid. We counted how many proposals were selected in both lists, how many moved into the final selection list and how many moved out of the final selection list due to the review meeting. We also looked at how many proposals didn't change rank, how many went up in rank and how many went down in rank, regardless of whether it changed their selection status. We looked at observing proposals and AR/TR proposals as two separate categories to determine if they were treated in any way systematically differently than the observing proposals.

The data for the GO-small observing proposals are provided in table 3. Column 1 identifies the panel, column 2 is the total number of GO-small proposals in the panel, columns 3-5 identify the percentage of proposals that retained the same rank, moved up in ranking and moved down in ranking at the review meeting, column 6 is the total number of GO-small proposals selected by the panel, column 7 is the percentage of proposals in the final selected list that would not have been selected based on the preliminary ranking and column 8 is the percentage of proposals in the preliminary list that did not make it into final selection. The same data for the AR/TR proposals is provided in table 4. We examine panel 1C in cycle-3 to illustrate the definition of the numbers in columns 7 and 8. Seventeen proposals were in the final list of

GO-small proposals selected by panel 1C in cycle-3. Four of those 17 selected proposals (24%) were not in the preliminary list. Their ranks moved up sufficiently to be in the selection. Three proposals from the preliminary list (18%) moved down in rank sufficiently to not make the final selection.

Table 3. The statistics for Cycles 3, 4 and 5 for how the GO-small proposal selection changed due to the changes in ranking of the proposals after discussions at the review meeting.

Small GO Proposals Cycle-3							
Panel	# Proposals in Panel	Proposal Ranking Changes			# Proposals Accepted	Moved into Selection	Moved out of Selection
		None	Up	Down			
1A	33	33%	6%	61%	11	0%	9%
1B	37	27%	43%	30%	11	27%	27%
1C	52	13%	54%	33%	17	24%	18%
2A	52	17%	46%	37%	17	35%	29%
2B	45	38%	36%	27%	13	15%	15%
3A	62	5%	47%	48%	17	24%	18%
3B	46	22%	50%	28%	11	36%	9%
3C	67	4%	42%	54%	30	37%	27%
4A	66	38%	32%	30%	19	32%	37%
4B	44	11%	52%	36%	14	29%	21%
5	27	7%	44%	48%	11	45%	36%
<b>Total</b>	<b>531</b>	<b>19%</b>	<b>42%</b>	<b>39%</b>	<b>171</b>	<b>29%</b>	<b>23%</b>

Small GO Proposals Cycle-4							
Panel	# Proposals in Panel	Proposal Ranking Changes			# Proposals Accepted	Moved into Selection	Moved out of Selection
		None	Up	Down			
1A	39	23%	44%	33%	11	27%	9%
1B	45	27%	49%	24%	13	38%	8%
1C	39	44%	26%	28%	9	33%	56%
2A	46	33%	30%	35%	9	33%	22%
2B	42	36%	21%	40%	10	10%	10%
3A	63	38%	29%	33%	11	45%	36%
3B	61	8%	34%	57%	17	24%	24%
3C	47	38%	23%	38%	12	33%	33%
4A	55	29%	44%	27%	17	24%	24%
4B	51	29%	33%	37%	14	21%	29%
5	36	28%	39%	33%	10	30%	10%
<b>Total</b>	<b>524</b>	<b>30%</b>	<b>34%</b>	<b>36%</b>	<b>133</b>	<b>29%</b>	<b>23%</b>

Small GO Proposals Cycle-5							
Panel	# Proposals in Panel	Proposal Ranking Changes			# Proposals Accepted	Moved into Selection	Moved out of Selection
		None	Up	Down			
1A	48	33%	38%	29%	9	33%	11%
1B	55	36%	20%	44%	13	31%	15%
1C	52	25%	35%	40%	11	45%	36%
2A	51	33%	41%	25%	8	38%	38%
2B	51	25%	41%	33%	11	27%	27%
3A	54	33%	35%	31%	10	40%	30%
3B	57	28%	40%	32%	11	45%	36%
3C	58	24%	43%	33%	14	29%	21%
4A	54	39%	28%	33%	12	17%	25%
4B	47	28%	40%	32%	11	27%	9%
5	29	52%	24%	24%	8	25%	25%
<b>Total</b>	<b>556</b>	<b>32%</b>	<b>35%</b>	<b>33%</b>	<b>118</b>	<b>32%</b>	<b>25%</b>

Table 4. The statistics for Cycles 3, 4 and 5 for how the Archival and Theoretical proposal selection changed due to the changes in ranking of the proposals after discussions at the review meeting.

AR/TR Proposals Cycle-3							
Panel	# Proposals in Panel	Proposal Ranking Changes			# Proposals Accepted	Moved into Selection	Moved out of Selection
		None	Up	Down			
1A	12	50%	25%	25%	3	33%	33%
1B	8	50%	38%	13%	1	0%	0%
1C	3	0%	0%	100%	1	0%	0%
2A	8	25%	38%	38%	2	50%	50%
2B	13	69%	15%	15%	1	100%	0%
3A	16	6%	31%	63%	3	0%	0%
3B	9	44%	44%	11%	3	0%	0%
3C	12	0%	33%	67%	4	75%	75%
4A	5	40%	20%	40%	1	0%	0%
4B	12	0%	33%	67%	5	20%	40%
5	4	25%	0%	75%	1	0%	300%
<b>Total</b>	<b>102</b>	<b>28%</b>	<b>28%</b>	<b>43%</b>	<b>25</b>	<b>28%</b>	<b>40%</b>

AR/TR Proposals Cycle-4							
Panel	# Proposals in Panel	Proposal Ranking Changes			# Proposals Accepted	Moved into Selection	Moved out of Selection
		None	Up	Down			
1A	4	0%	50%	50%	1	0%	0%
1B	5	40%	40%	20%	2	100%	50%
1C	12	33%	58%	8%	4	50%	0%
2A	7	29%	43%	29%	3	0%	0%
2B	5	40%	40%	20%	2	0%	0%
3A	16	50%	38%	13%	6	33%	17%
3B	10	0%	60%	40%	2	50%	0%
3C	14	14%	43%	43%	7	14%	0%
4A	13	54%	38%	8%	5	80%	0%
4B	9	33%	44%	22%	2	100%	0%
5	6	50%	17%	17%	2	0%	0%
<b>Total</b>	<b>101</b>	<b>33%</b>	<b>44%</b>	<b>23%</b>	<b>36</b>	<b>39%</b>	<b>6%</b>

AR/TR Proposals Cycle-5							
Panel	# Proposals in Panel	Proposal Ranking Changes			# Proposals Accepted	Moved into Selection	Moved out of Selection
		None	Up	Down			
1A	8	25%	38%	38%	2	50%	100%
1B	7	29%	14%	57%	3	0%	0%
1C	6	17%	33%	50%	3	33%	0%
2A	7	57%	14%	29%	2	0%	0%
2B	16	19%	50%	31%	5	40%	0%
3A	13	15%	54%	31%	3	0%	0%
3B	11	18%	45%	36%	5	60%	20%
3C	14	50%	36%	14%	4	25%	0%
4A	12	33%	42%	25%	6	17%	17%
4B	11	36%	18%	45%	3	0%	0%
5	4	25%	25%	50%	1	0%	100%
<b>Total</b>	<b>109</b>	<b>29%</b>	<b>37%</b>	<b>34%</b>	<b>37</b>	<b>24%</b>	<b>14%</b>

The panel meeting, discussion and final grading of the proposals have a substantial impact on the make-up of the final ranked list of proposals. Approximately 30% of the proposals selected during cycles 3, 4 and 5 would not have been awarded time if only the preliminary grades were utilized in the selection. Approximately 25% of the proposals that

would have been selected using the preliminary grades do not make it into the final ranked list. The percentage of proposals moving out of and into the final ranked list do not match because at the margins smaller programs are more likely to move up and larger programs move down the list. Though the call for proposals and instructions to the review panels specifically discourage cutting the time requested in a proposal the review panels do recommend awarding proposals less time for about 5% of the proposals. This also leads to a slightly larger number being accepted than would be the case if no cuts were made.

There appears to be more movement into and out of the final selection list for the observing proposals compared to the AR/TR proposals. For all three cycles studied, four to five panels/cycle made no change from the preliminary list to the final list for selected AR/TR programs. There are no cases of panels that didn't change the final selected list based on the review discussions for the observing proposals. There are cases of substantial change for the AR/TR programs, i.e. panel 5 in cycle-3 selected 1 AR/TR proposal in the final list though there were 4 in the preliminary list. This is displayed graphically in figures 4 and 5. Figure 4 shows the percentage of proposals that moved into and out of the final selection list for the observing proposals and figure 5 shows the same information for the AR/TR proposals.

Fig. 4. The percentage of the proposals that moved into and out of the final selection list, based on the discussion at the review meeting, for the GO-small observing proposals. This shows graphically the data in columns 7 and 8 of table 3.

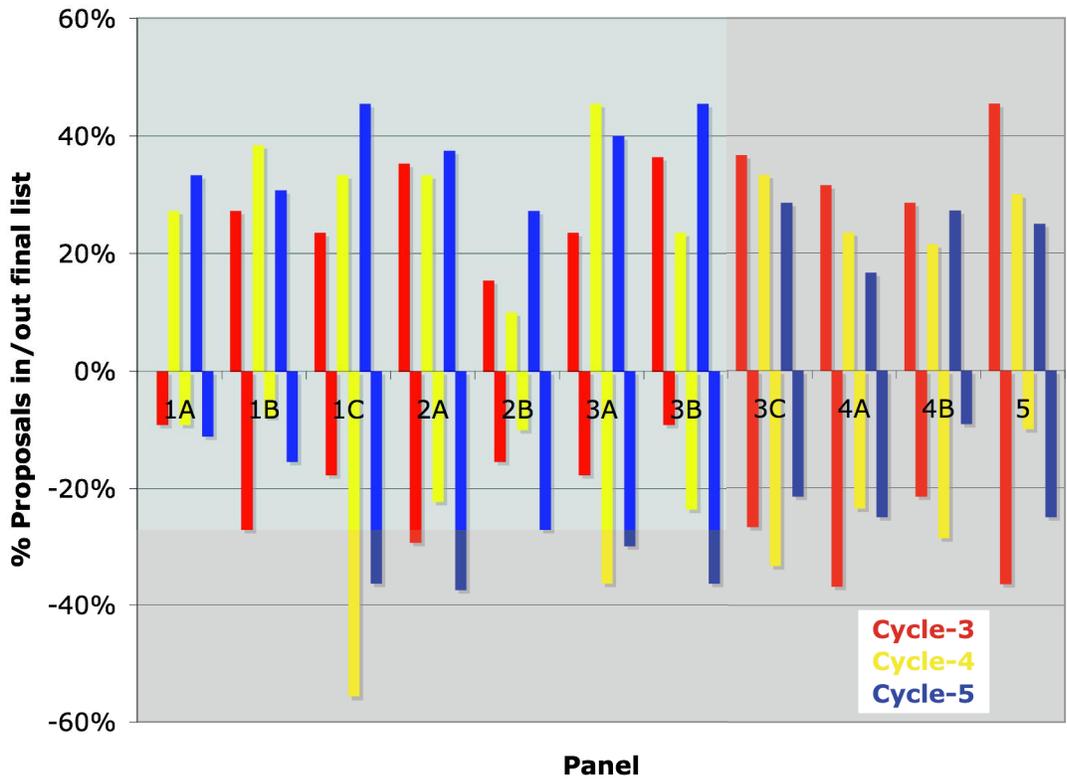
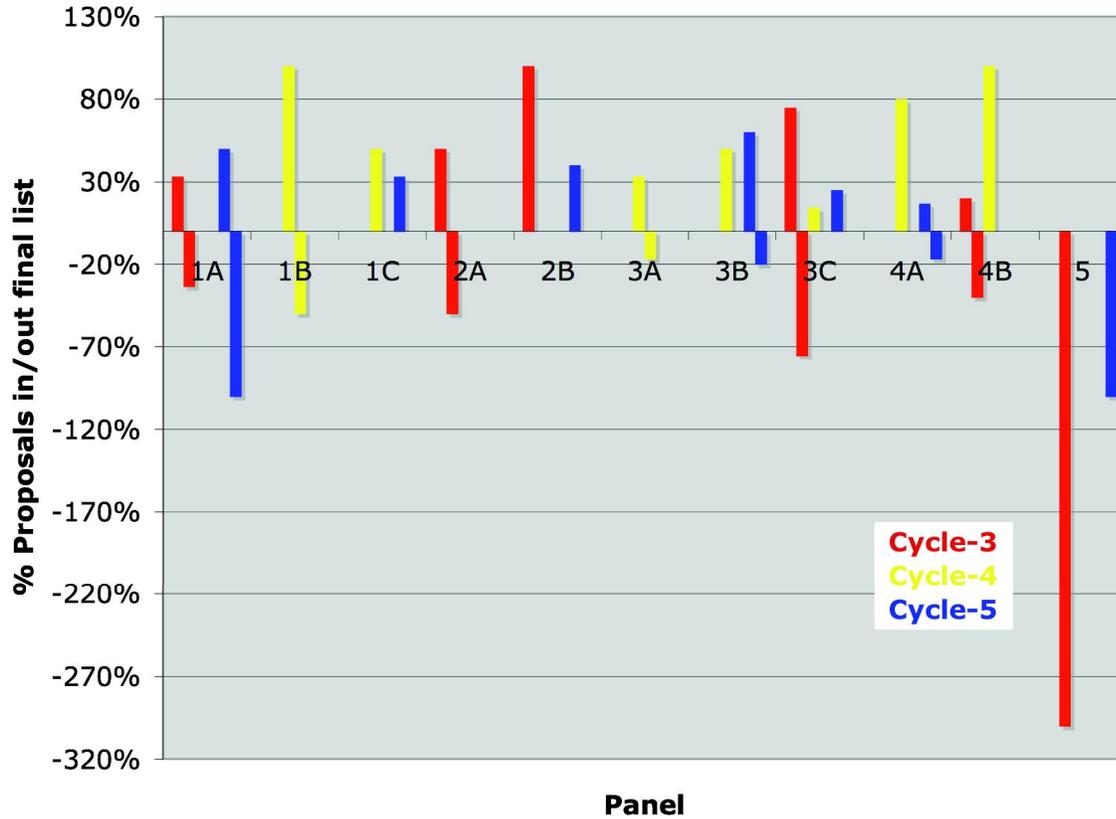


Fig. 5. The percentage of the proposals that moved into and out of the final selection list, based on the discussion at the review meeting, for the AR/TR proposals. This shows graphically the data in columns 7 and 8 of table 4.



#### 4. CONCLUSIONS

Based on the impact the discussions at the review meeting have consistently had on the final ranked list of observing proposals it would be imprudent for the SSC to go to a completely remote proposal review process, i.e. just obtaining grades from the reviewers without allowing for any form of panel discussion. If the effect of the panel discussions changed the final answer by 5-10% it would probably be reasonable to just use the preliminary grades. Since the effect is 25-35% we would not be comfortable with this model. The options we can consider are virtual meetings (via telecon, ichtat or Skype) or exploring the NRAO model that instead of having broad-based panels has expert referees that only review proposals in their field. The results from the expert referees are forwarded to a ‘merging’ TAC that reviews the results from the referees and determines the final ranking for any proposals where the referees disagree. If the Spitzer Warm Mission is approved we will formulate a plan for the reviews and consult with our Oversight Committee and User’s Panel before implementing the new plan. Spitzer plans to continue using face-to-face meetings for the selection of large programs.

#### REFERENCES

- [1] Laine, S., Silbermann, N.A., Rebull, L.M. and Storrie-Lombardi, L.J., “Spitzer Space Telescope proposal process,” Proc. SPIE 6270, 29 (2006).
- [2] Crane, M., Storrie-Lombardi, L.J., Silbermann, N.A., Rebull, L.M., “SPRITE: the Spitzer proposal review website,” to appear in Proc. SPIE 7016 (2008)