The Influence of Peer Reviewer Expertise on the Evaluation of Research Funding Applications

Stephen Gallo, PhD

Technical Operations Manager

American Institute of Biological Sciences

Scientific Peer Advisory and Review Services

American Institute of Biological Sciences

Studying Reviewer Decision-Making

- *"NIH peer review percentile scores are poorly predictive of grant productivity" Fang et al., 2016*

- Decision-making processes understudied

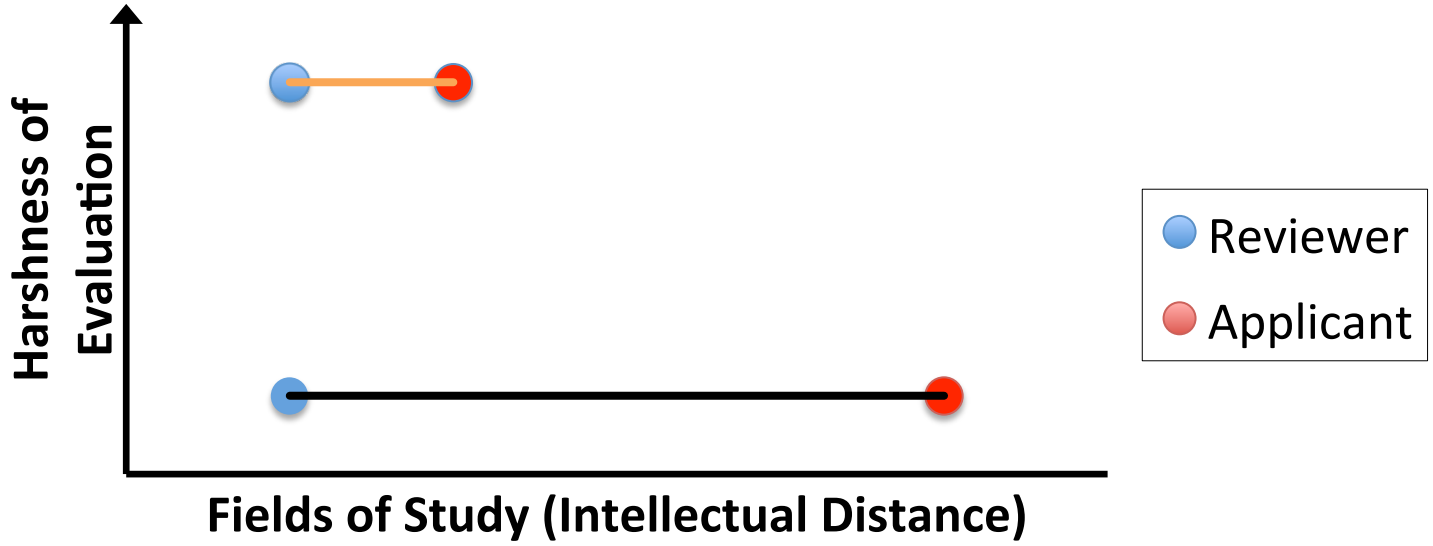
- What drives reviewers' evaluations?

Reviewer Decision-Making Model

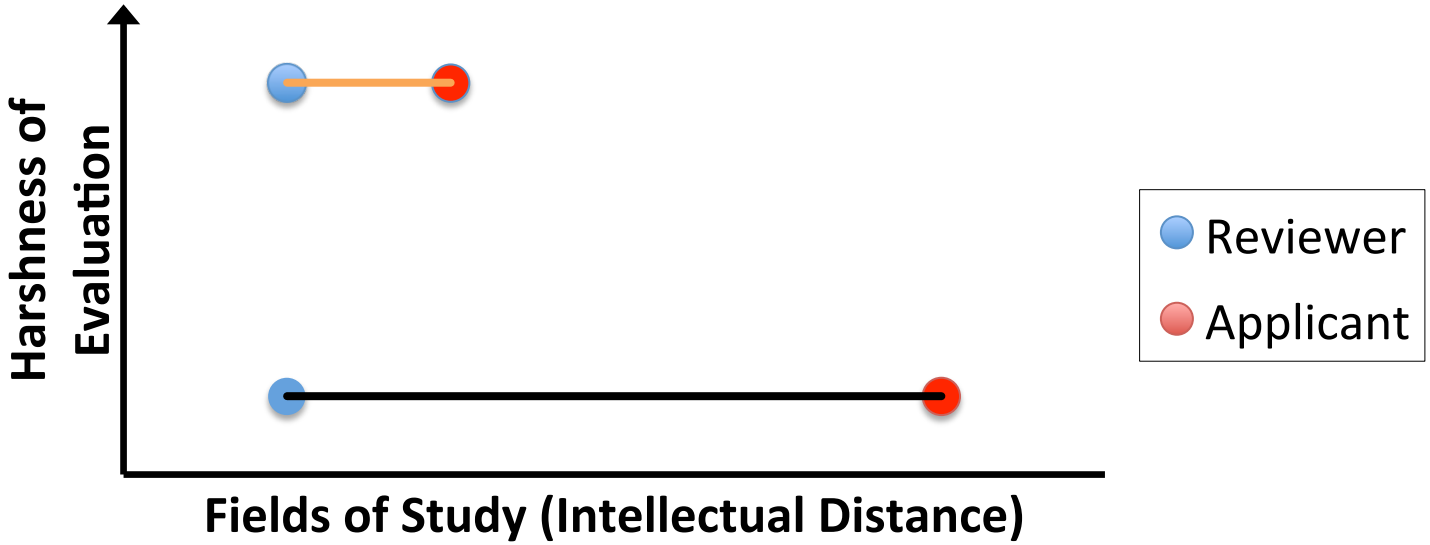
- Boudreau (2016) study used ***blinded***, individual reviews of grant applications
- Found shorter intellectual distance (***keyword based***) between applicant and reviewer leads to harsher evaluations

Reviewer Decision-Making Model

- Boudreau (2016) study used ***blinded***, individual reviews of grant applications
- Found shorter intellectual distance (***keyword based***) between applicant and reviewer leads to harsher evaluations



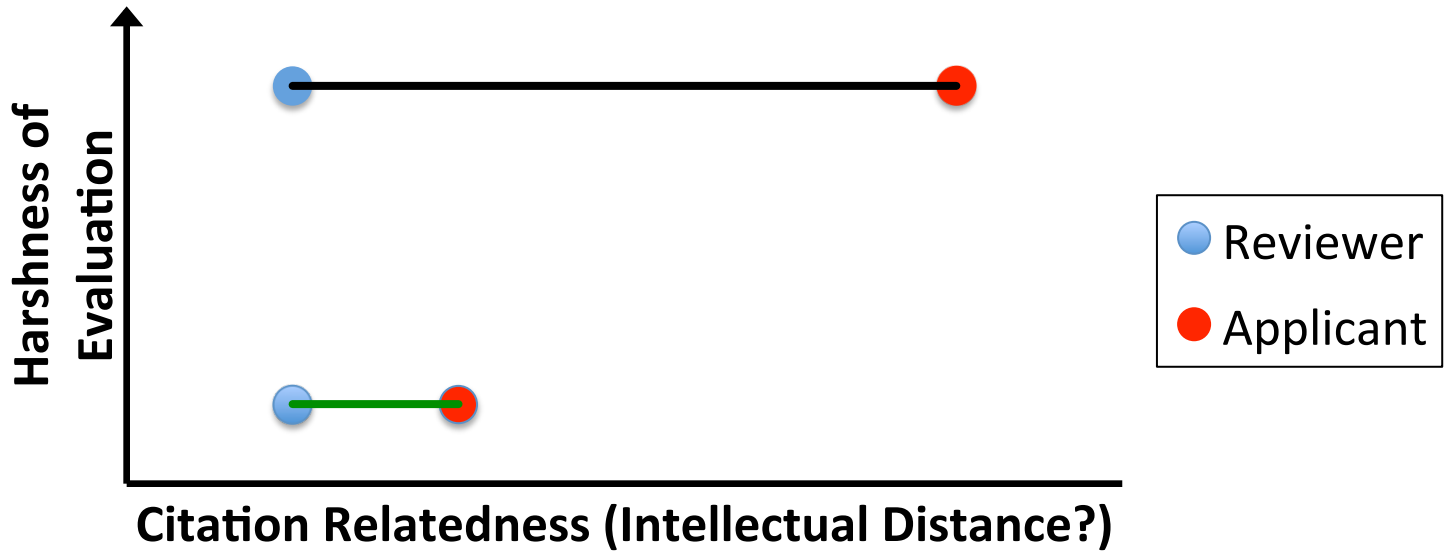
Reviewer Decision-Making Model



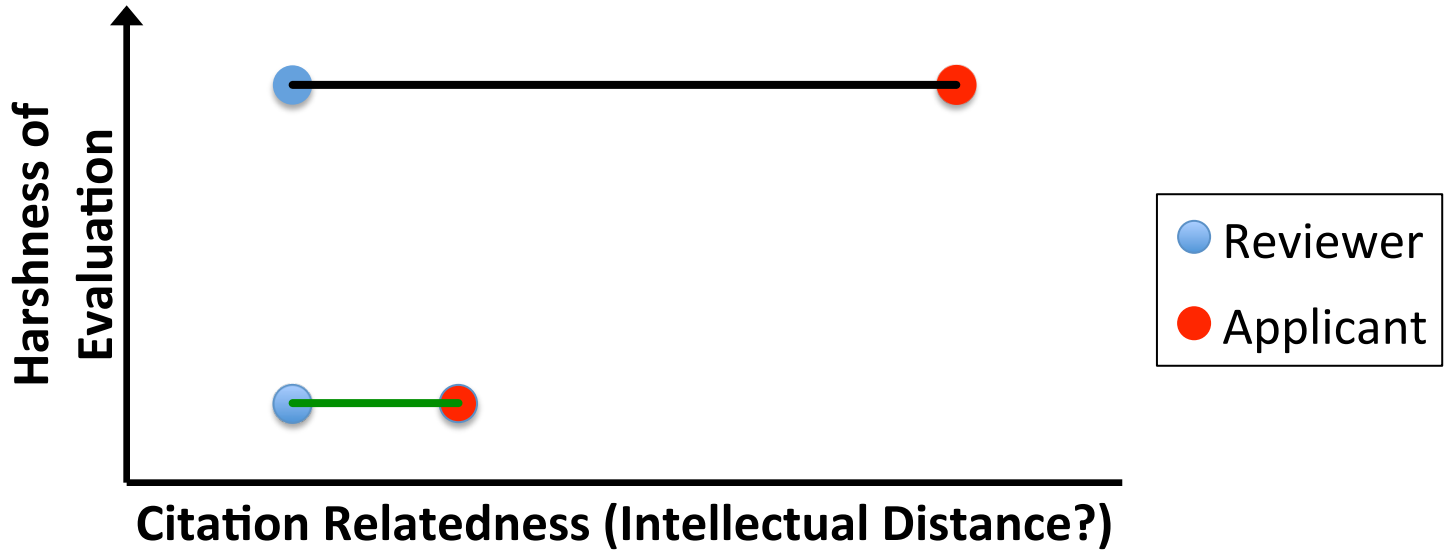
- Information sampling effect: higher expertise (likely shorter intellectual distance), more information sampled
- Increased sampling is skewed toward weaknesses; harsh evaluations

Contrasting Results

- Li (2013) study of **un-blinded** NIH review panels
- Found higher degree of relatedness (***citation linkage***) between applicant and reviewer yields more favorable evaluations



Contrasting Results



- More "related" reviewers sample additional information and make inferences about application quality (therefore may be less harsh)
- Relatedness may be indirect measure of intellectual distance?

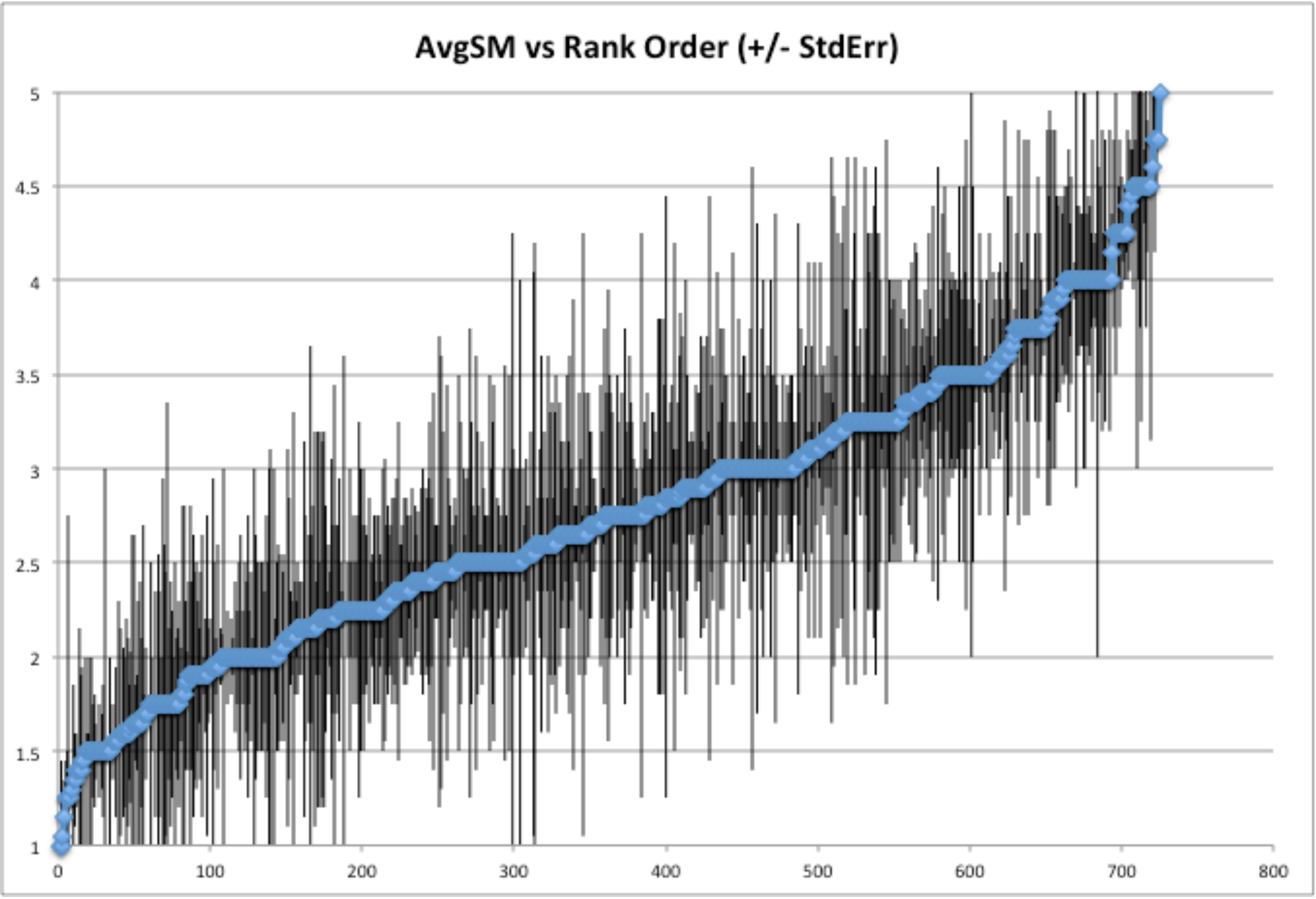
AIBS Hypothesis

- AIBS data: scoring (**scientific merit [SM]** and self-assessed **reviewer expertise [RE]**)
- SM vs. RE: **higher expertise, harsher evaluations/scoring**
- However, in **un-blinded** scenario, relationship will be **sensitive to reviewer and applicant seniority**

Retrospective Analysis of AIBS Peer Review Data

- **Un-blinded** peer review program of biomedical research applications (anonymized scoring/commentary)
- 1,450 individual reviews of 725 applications (2009 to 2012)
- **SM** scale: 1 – 5; 1 is highest merit
- **RE** scale: 1 – 5; 1 is highest expertise
- Examine relationship between **SM** and **RE**

Reviewer Demographics (Total Reviewers = 1044)			Applicant Demographics (Total Proposals = 725)	
	N	**%**	**N**	**%**
Gender				
M	800	77	619	85
F	244	23	106	15
Position				
Junior	277	27	130	18
Non-Junior	767	73	595	82
Sector				
Acad	961	92	429	59
Non_Acad	83	8	296	41
Degree				
No MD	718	69	476	66
MD Training	326	31	249	34


AvgSM vs Rank Order (+/- StdErr)
5
4.5
4
3.5
3
2.5
2
1.5
1
0
100
200
300
400
500
600
700
800

Analytic Approach

- Hierarchical data
- Multi-level multiple regression approach
- Individual reviewer evaluations (level 1) nested in application groupings (level 2)
- Random intercept model:

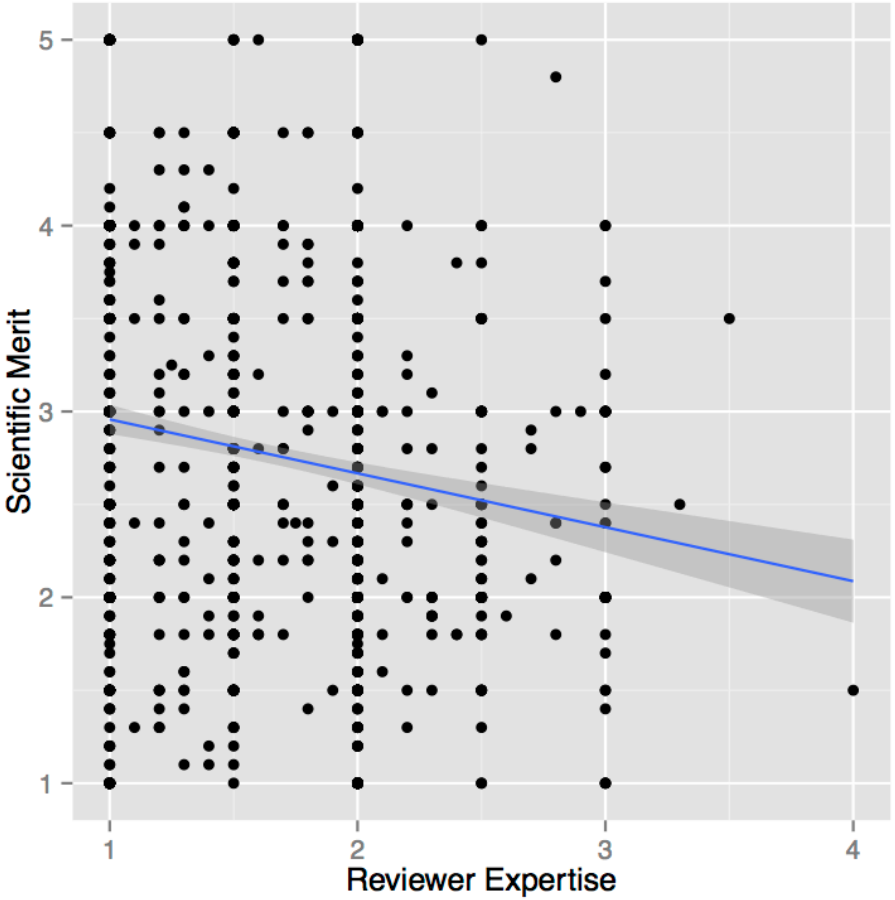
$$\mathbf{Y_{ij} = X_{hij}\beta_h + \beta_0 + \mu_{0j} + \varepsilon_{ij}}$$

Y_{ij} = scientific merit of reviewer *i* of the *j*th application
X_{hij} = vector of *h* independent variables with *h* coefficients β_h
β_0 = constant intercept
μ_{0j} = random intercept varies across applications
ε_{ij} = residual error component

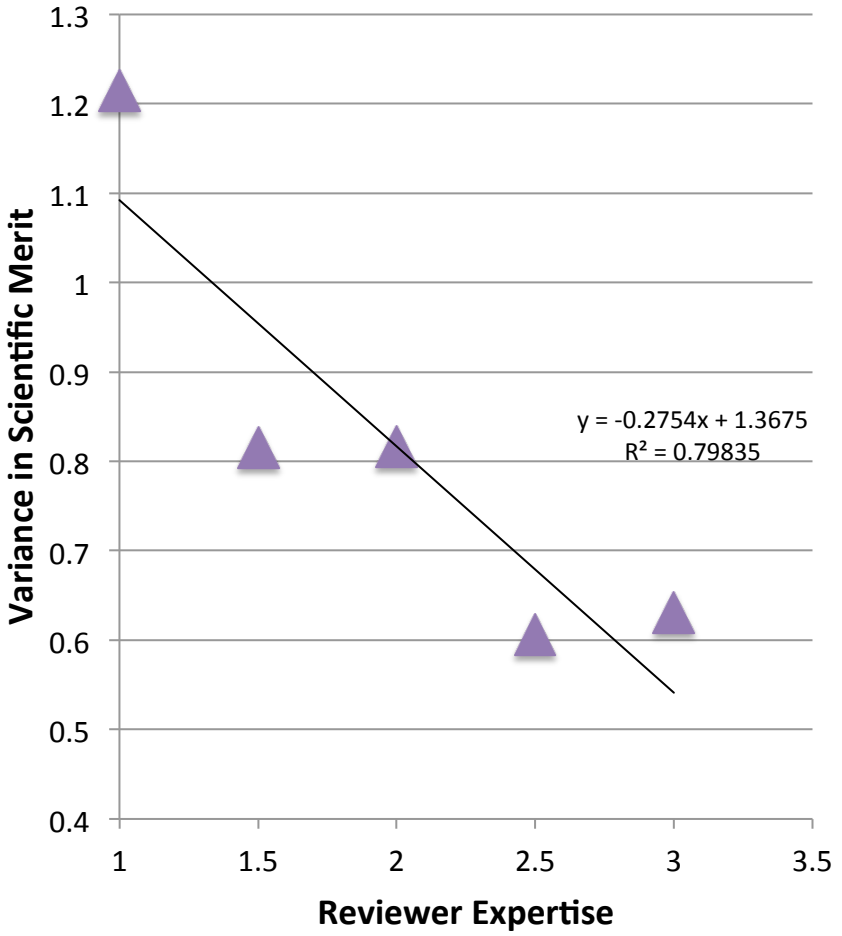
Effect Type	Model 1: Baseline across applications	Model 2: RE	Model 3: RE + Reviewer Demographics	Model 4: RE + Reviewer and Applicant Demographics	Model 5: RE + Reviewer Demographics + Applicant Position
Random Effects					
Variance Across Proposals	0.219 (0.042)*	0.213 (0.042)*	0.204 (0.042)*	0.197 (0.042)*	0.201 (0.043)*
Variance Across Assesors	0.740 (0.023)*	0.722 (0.023)*	0.716 (0.023)*	0.713 (0.023)*	0.714 (0.023)*
Fixed Effects					
Intercept	2.77 (0.03)*	3.24 (0.08)*	3.13 (0.12)*	3.07 (0.17)*	3.04 (0.13)*
Reviewer Expertise	_____	-0.28 (0.05)*	-0.19 (0.07)*	-0.20 (0.10)*	-0.16 (0.07)*
Deviance (previous model)					
Change in 2LL	38.9*	36.8*	21.5*	14.0	7.3* (compared to Model 3)

* Indicates statistical significance $p<0.05$

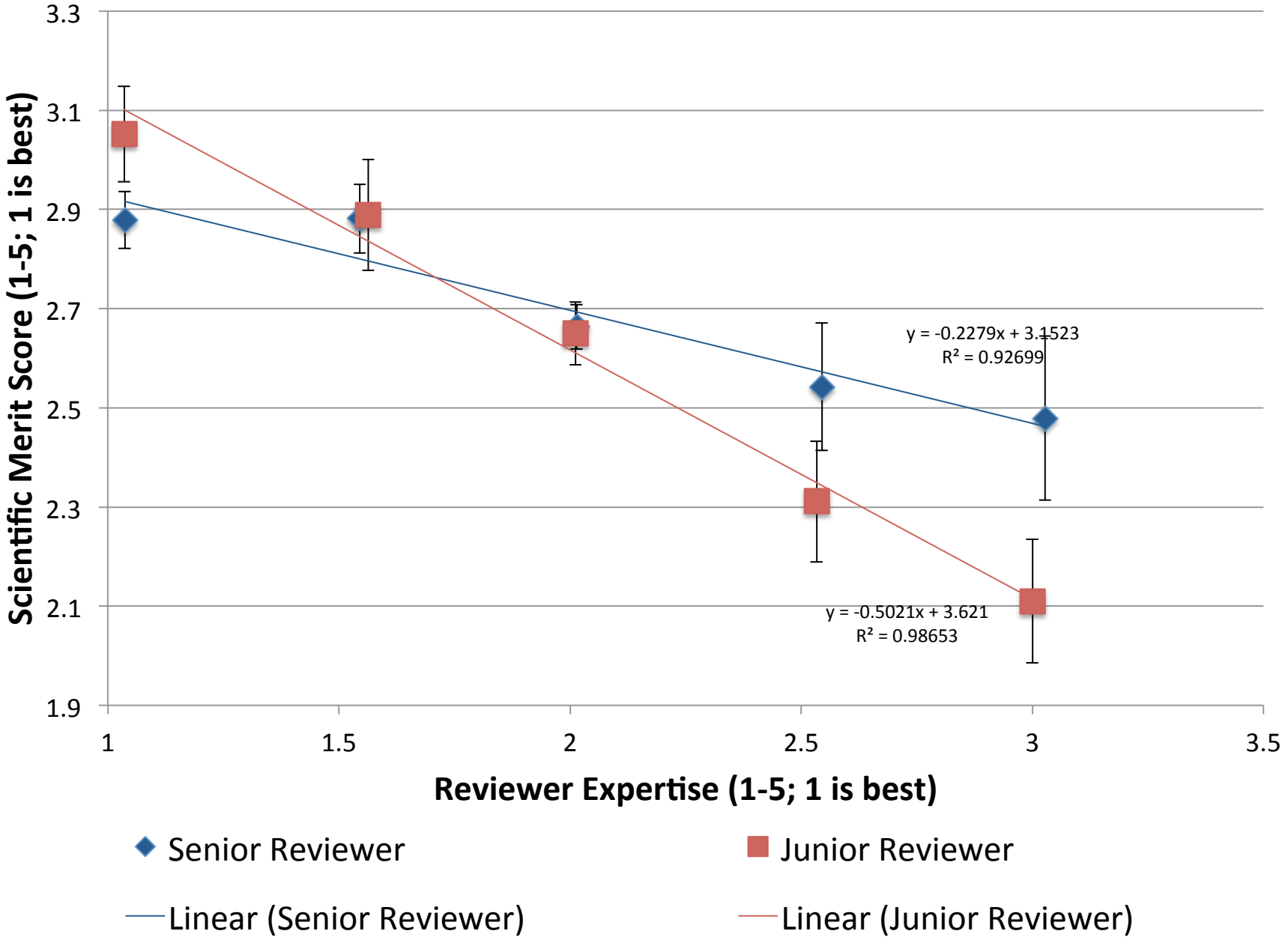
SM vs. RE



SM variance vs. RE



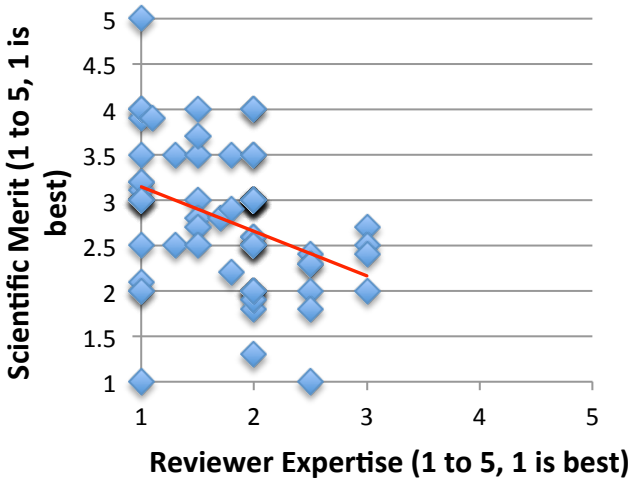
Model 5: RE + Reviewer Demographics + Applicant Position	
Random Effects	
Variance Across Proposals	0.201 (0.043)*
Variance Across Reviewers	0.714 (0.023)*
Intra-proposal Correlation	0.22*
Inter-Rater Reliability	0.36
Main Effects	
Intercept	3.04 (0.13)*
Reviewer Expertise (RE)	**-0.16 (0.07)***
Reviewer Gender (RevF)	-0.02 (0.20)
Junior Reviewer (RevJ)	**0.41 (0.19)***
Non-Academic Reviewer (RevNonAc)	-0.21 (0.29)
Reviewer with MD Degree (RevMD)	0.08 (0.17)
Junior Applicant (PIJ)	**0.48 (0.22)***
Interaction Effects	
Reviewer Expertise * Reviewer Gender	-0.03 (0.11)
Reviewer Expertise * Junior Reviewer	**-0.22 (0.11)***
Reviewer Expertise * Non-Academic Reviewer	-0.06 (0.17)
Reviewer Expertise * Reviewer with MD Degree	-0.07 (0.10)
Reviewer Expertise * Junior Applicant	-0.19 (0.12)

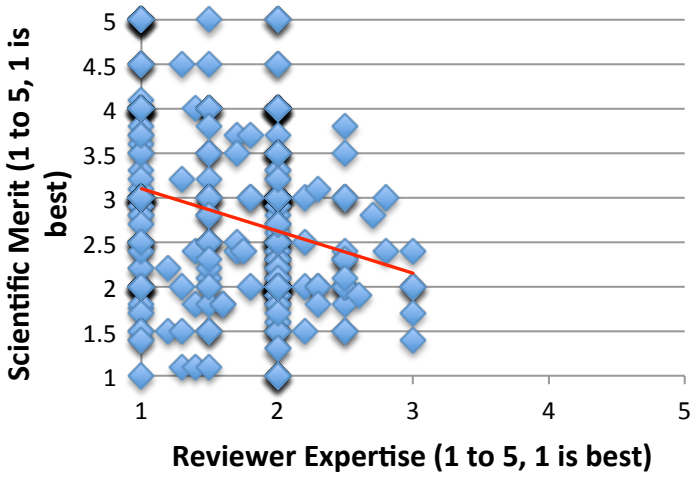

Scientific Merit Score (1-5; 1 is best)
3.3
3.1
2.9
2.7
2.5
2.3
2.1
1.9
1
1.5
2
2.5
3
3.5
Reviewer Expertise (1-5; 1 is best)
y = -0.2279x + 3.1523
R² = 0.92699
y = -0.5021x + 3.621
R² = 0.98653
Senior Reviewer
Junior Reviewer
Linear (Senior Reviewer)
Linear (Junior Reviewer)

Junior Applicant

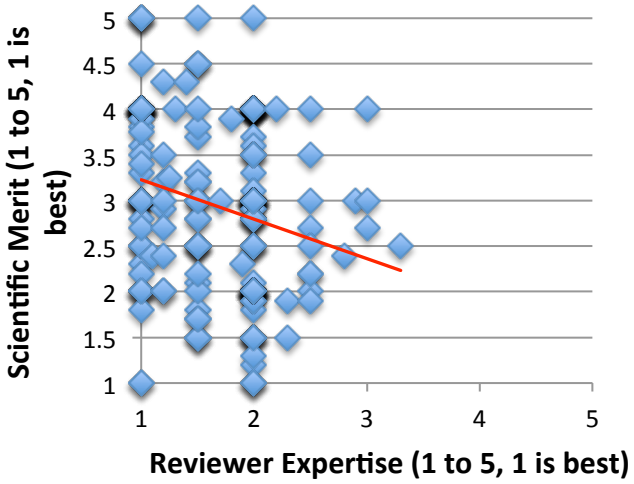
Non-Junior Applicant

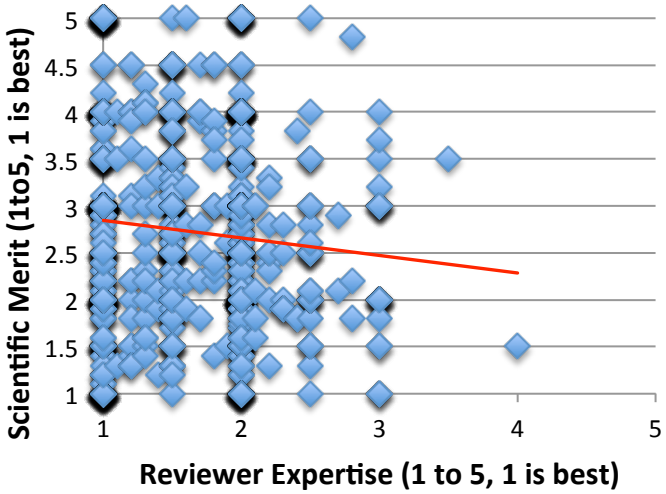
Junior Reviewer





Non-Junior Reviewer





Conclusions I

- We found **RE linearly correlated to SM** in **un-blinded** review
- Found **shorter intellectual distance** leads to **harsher evaluations**: information sampling
- However, SM vs. RE relationship **sensitive to applicant and reviewer seniority**

Conclusions II

- Senior reviewers may heavily weigh applicant track record
- Track record may compensate for any perceived methodological hurdles
- Intuition

Conclusions III

- Further research:
  - What types of information reviewers are sampling
  - How precisely reviewers are weighting this information

American Institute of Biological Sciences