Fairness and Bias in Peer Review and other Sociotechnical Intelligent Systems

Nihar B. Shah and Zachary Lipton

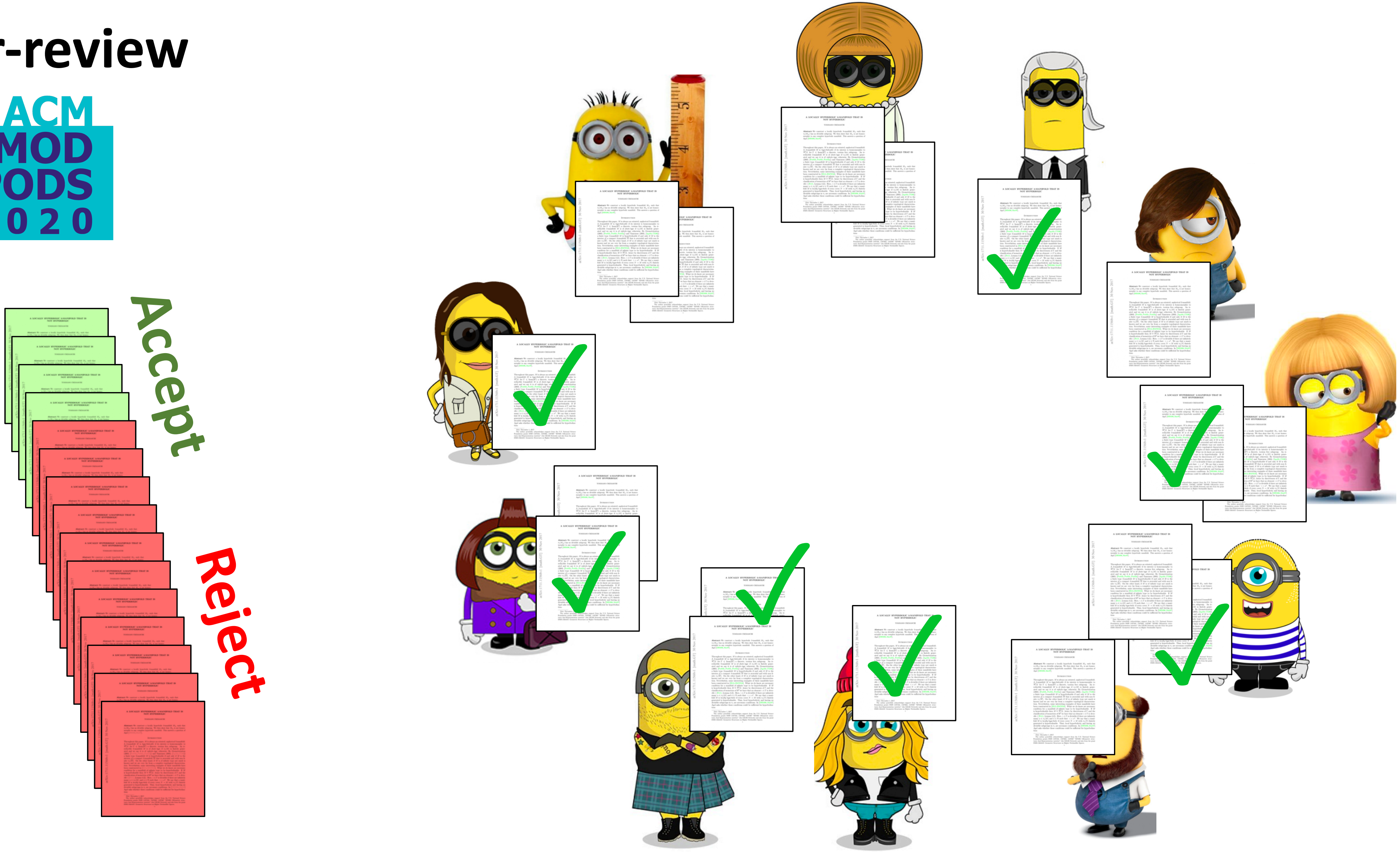
School of Computer Science

Carnegie Mellon University

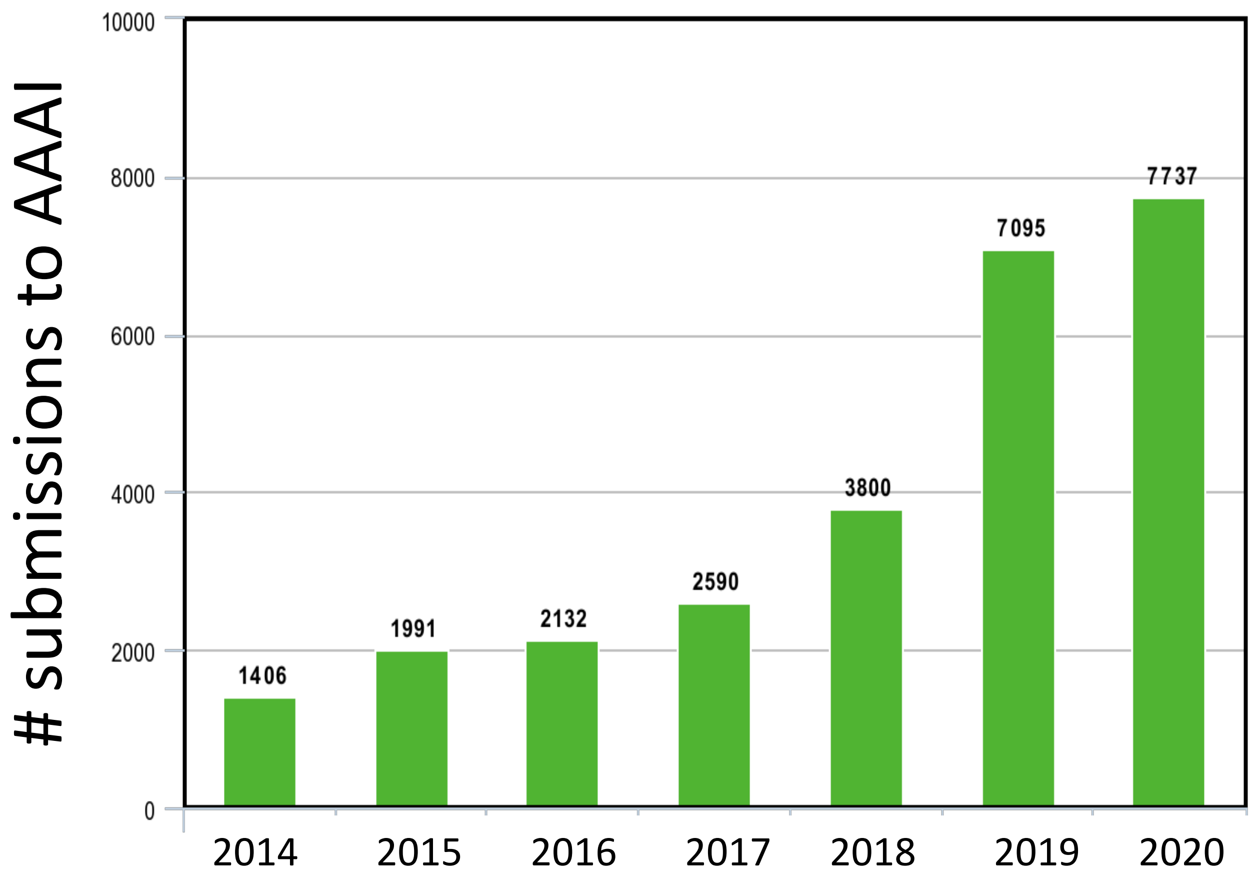
"Piled Higher and Deeper" by Jorge Cham WWW.PHDCOMICS.COM

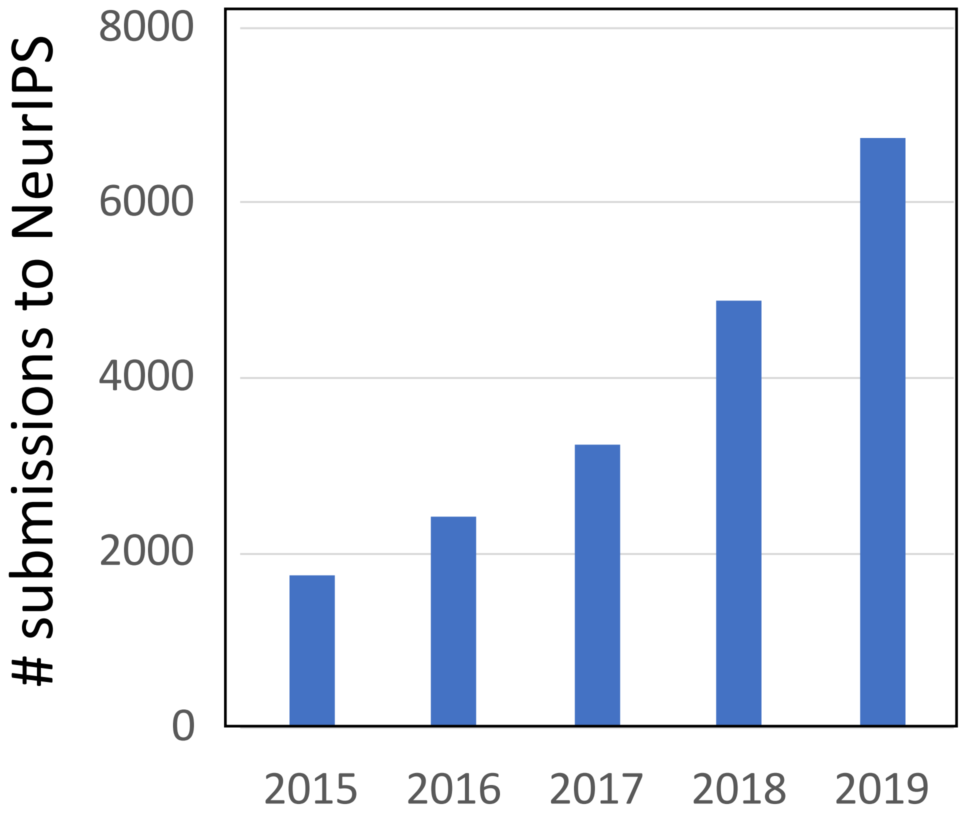
Peer-review

ACM SIGMOD + PODS 2020



Several thousand submissions, exponential growth



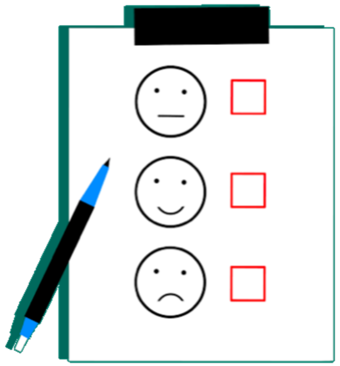


Challenge across many research fields

- **"Let's make peer review scientific"** [Rennie, Nature 2016]

  "Peer review … is a human system. Everybody involved brings ***prejudices****,* ***misunderstandings*** *and gaps in knowledge, so no one should be surprised that peer review is often* ***biased*** *and* ***inefficient****. It is occasionally* ***corrupt****, sometimes a charade, an open temptation to plagiarists. Even with the best of intentions, how and whether peer review identifies high-quality science is unknown. It is, in short,* ***unscientific****."*

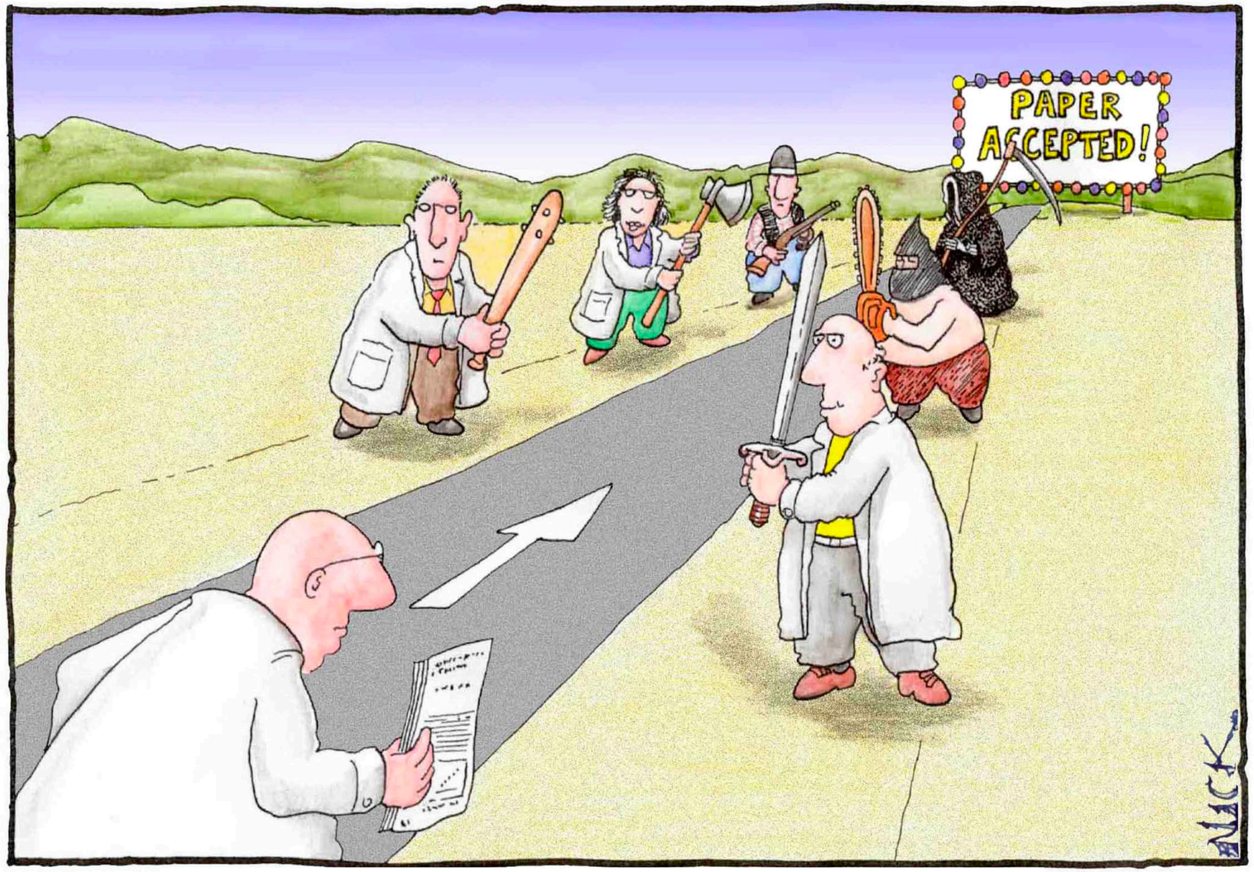
- **Overwhelming desire for improvement**
  [surveys by Smith 2006, Ware 2008, Mulligan et al. 2013]



Hurts scientific progress

"interdisciplinary research, frontier science, areas of controversy, and risky new departures are all more likely to suffer from cognitive cronyism than is mainstream research" [Travis et al. 1991]

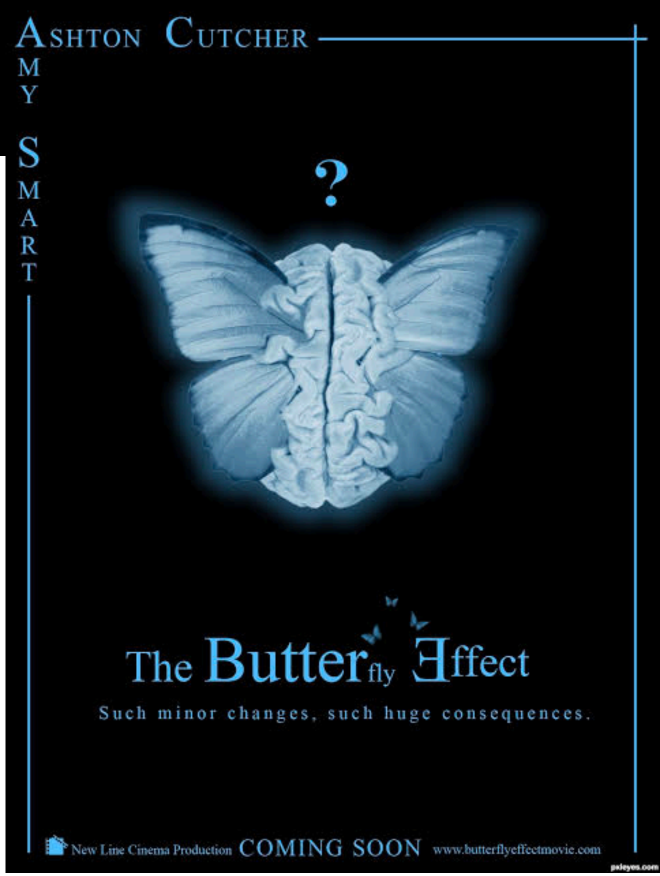
"Reviewers love safe (boring) papers, ideally on a topic that has been discussed before (ad nauseam)...The process discourages growth" [Church 2005]



Hurts careers

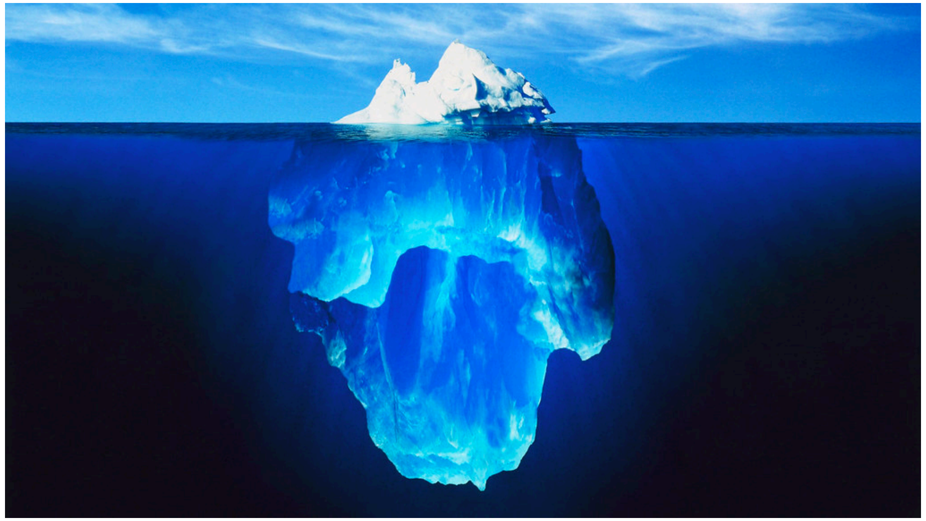
“an incompetent review may lead to the rejection of the submitted paper, or of the grant application, and the ultimate failure of the career of the author.” [Triggle et al. 2007]

“These long term effects arise due to the widespread prevalence of the Matthew effect (‘rich get richer’) in academia” [Merton 1968]



What are some of the challenges and what can be done about them?

This talk outlines some research being done.



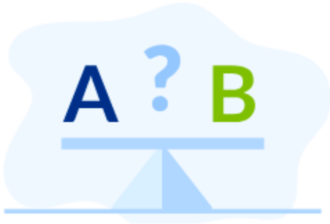
Calls upon YOUR expertise!

Broad applicability

Distributed human evaluations

Hiring

Admissions



A/B testing

Crowdsourcing

tripadvisor yelp

Product ratings

Healthcare

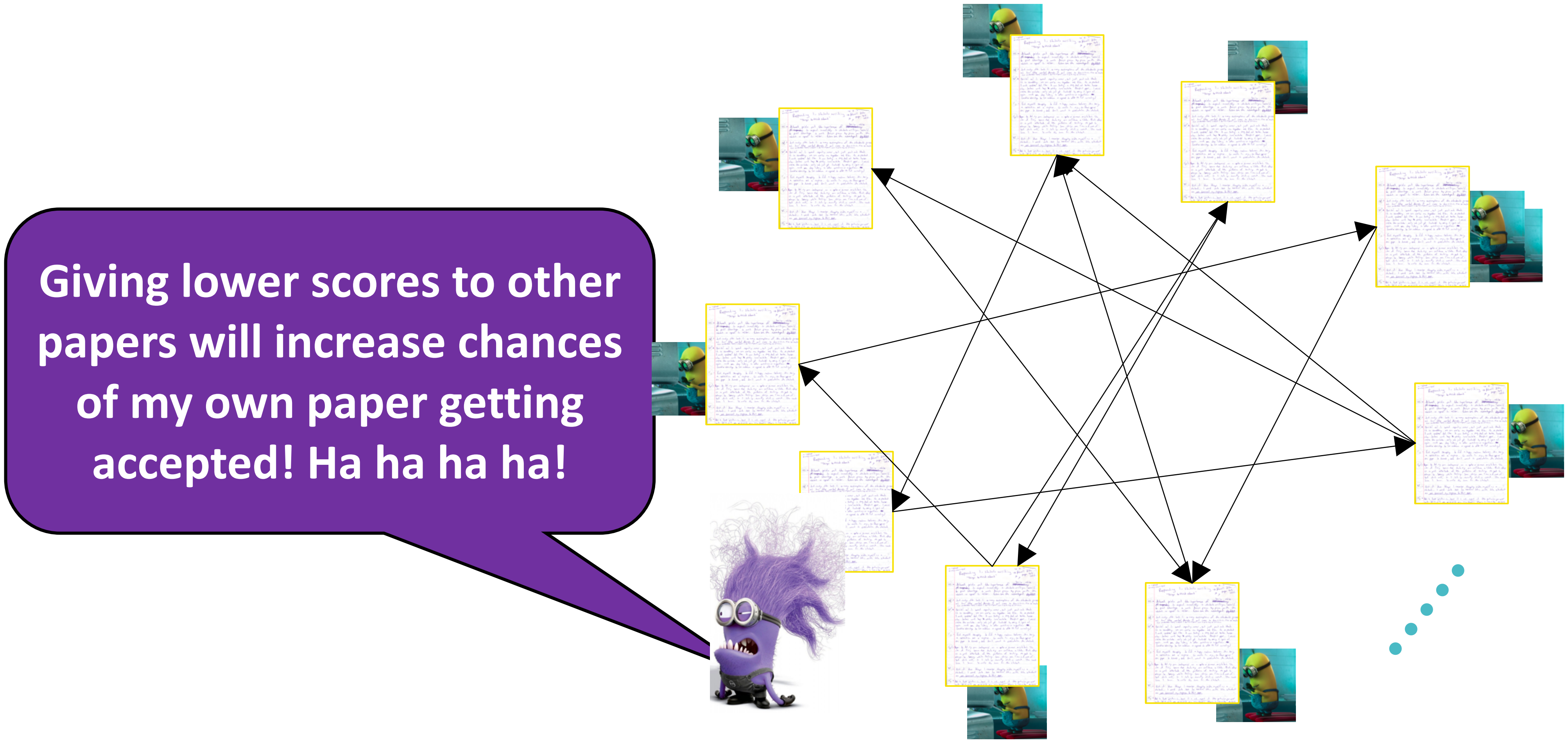
Peer grading

...

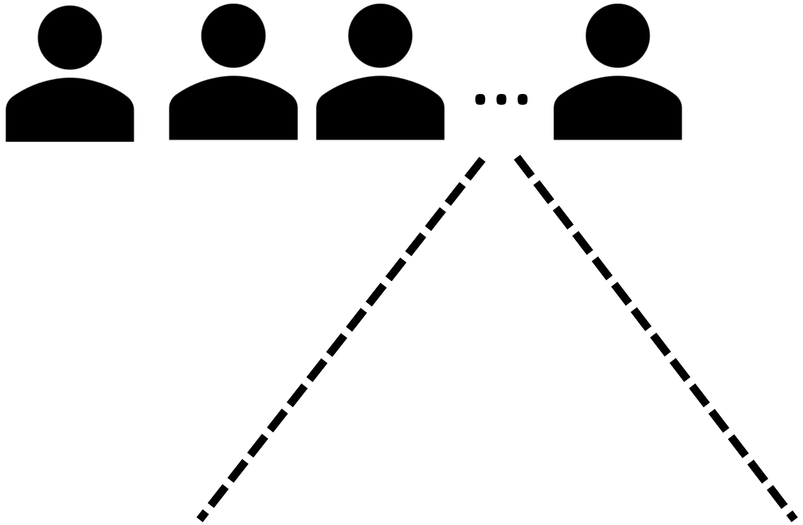
Problems amplify when this data is used to train AI/ML systems!

- Dishonest behavior
- Noise
- Miscalibration
- Subjectivity
- Biases
- Norms and policies

Dishonest behavior



An experiment



1. Make a drawing
2. Enter one of 3 “exhibitions”
3. Peer review others’ drawings
4. Possibly win an award

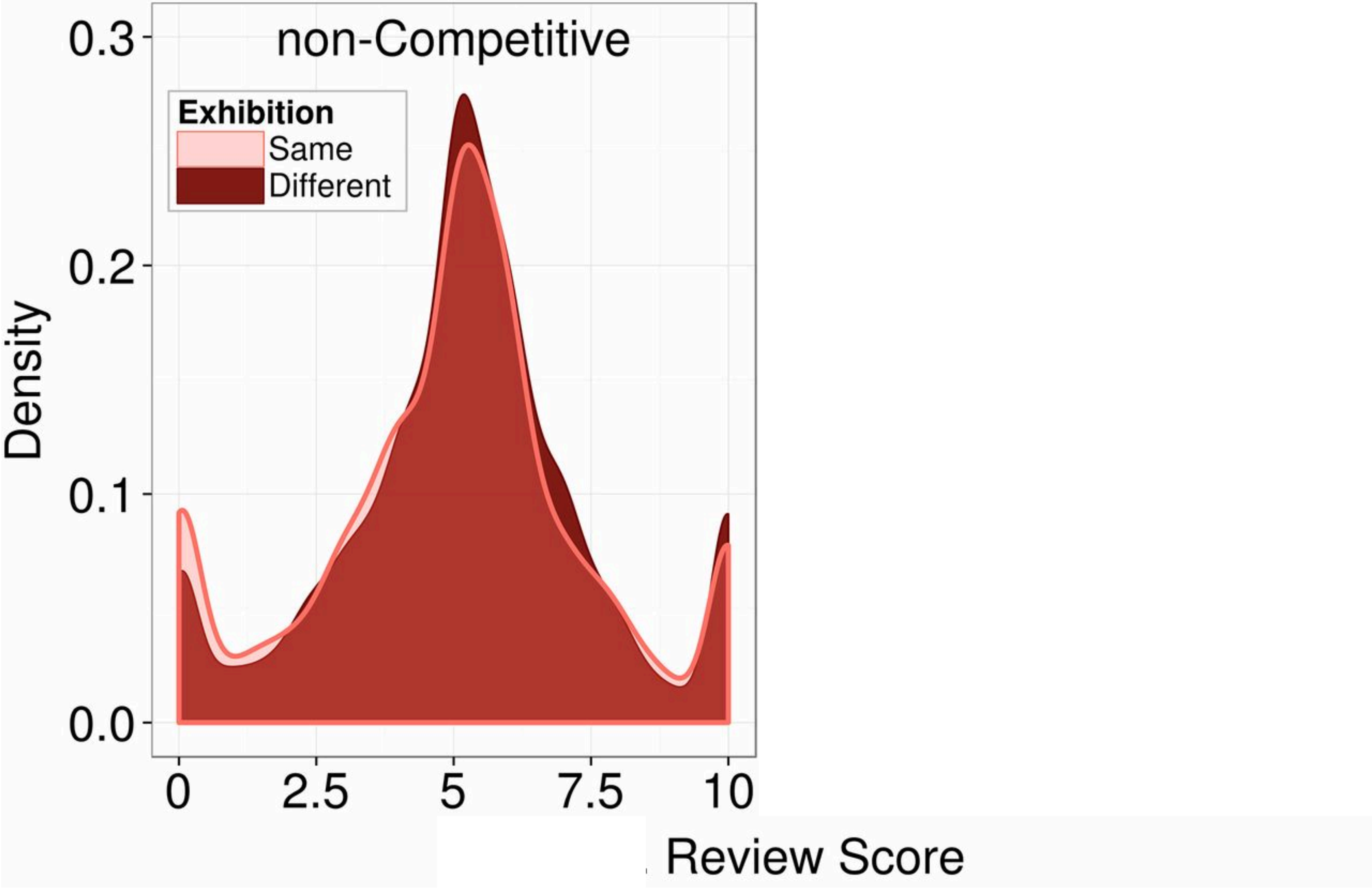
Non-competitive

All above certain threshold get award

Competitive

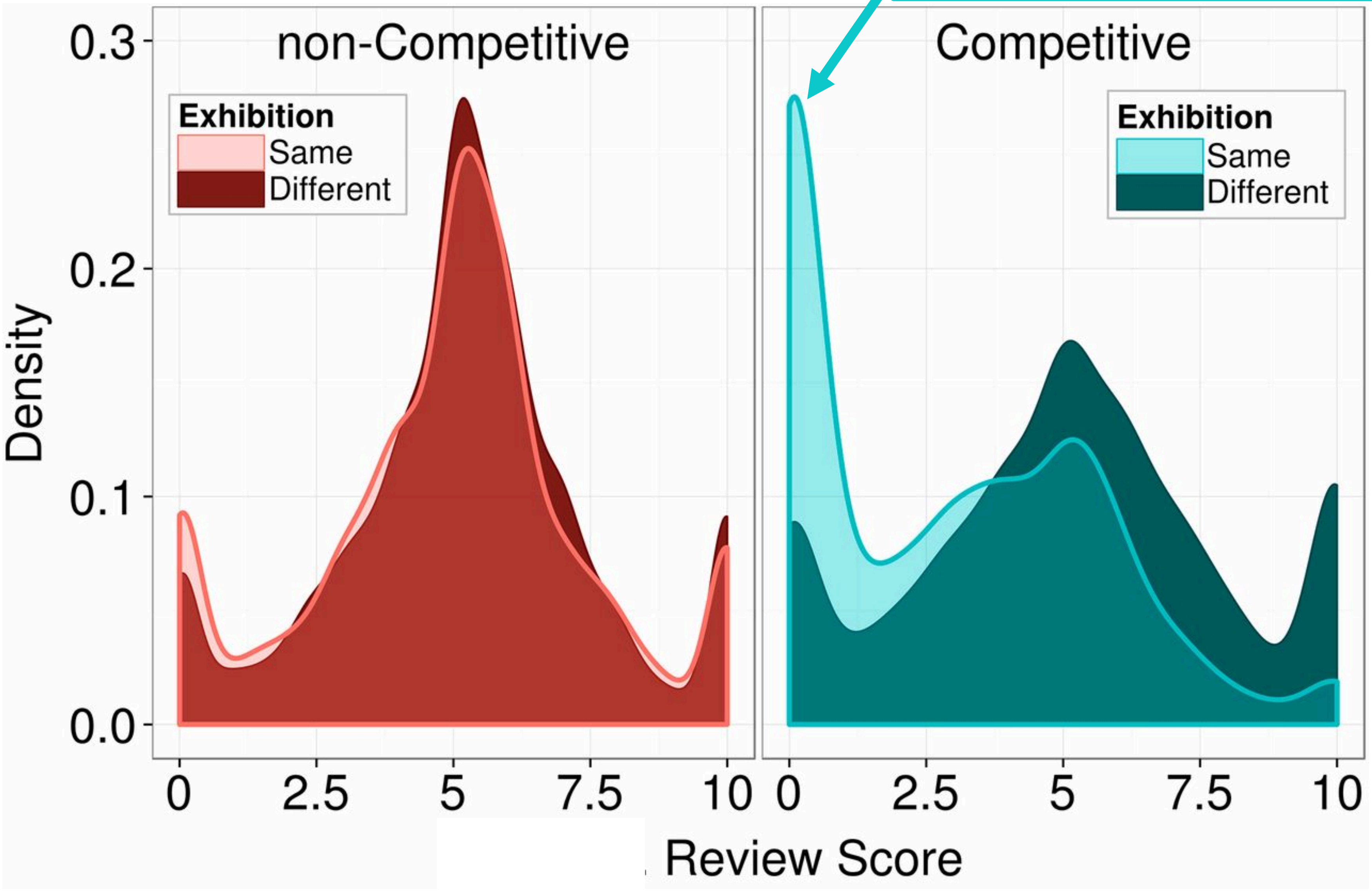
Top certain fraction in each exhibition win award

[Balietti et al., 2016]

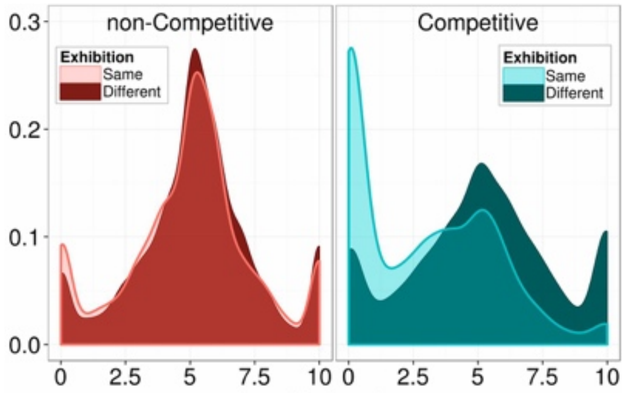


[Balietti et al., 2016]

Giving a lower score increases chances of their drawing getting an award



[Balietti et al., 2016]



- “competitive sessions produce considerably more [strategic] reviews”
- “the number of [strategic] reviews increases over time”

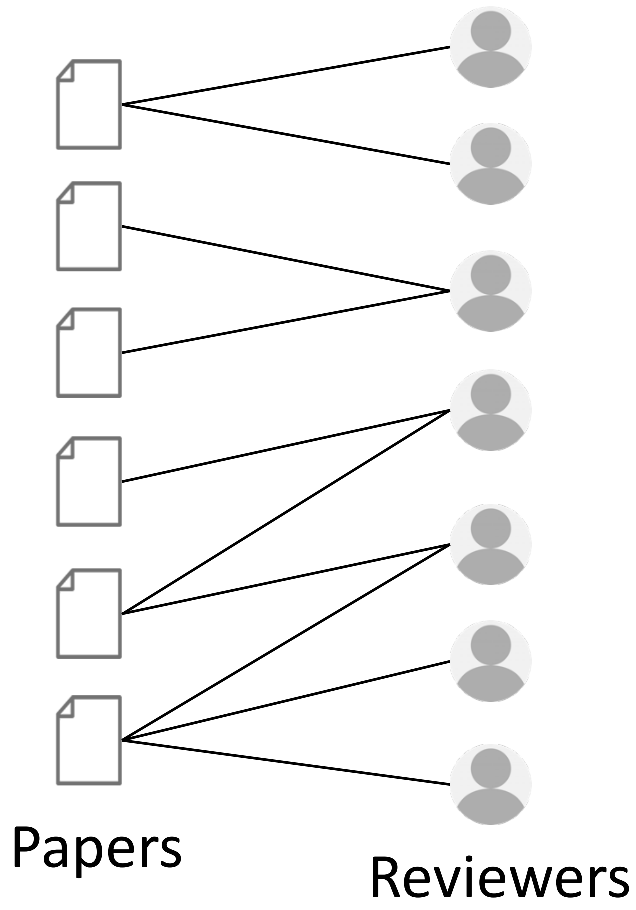
“This result provides further evidence that a substantial amount of gaming of the review system is taking place… competition incentivizes reviewers to behave strategically, which reduces the fairness of evaluations”

[Balietti et al., 2016]

Also [Anderson et al. 2007, Langford 2008 (blog), Akst 2010, Thurner and Hanel 2011]

How to make peer review strategyproof?

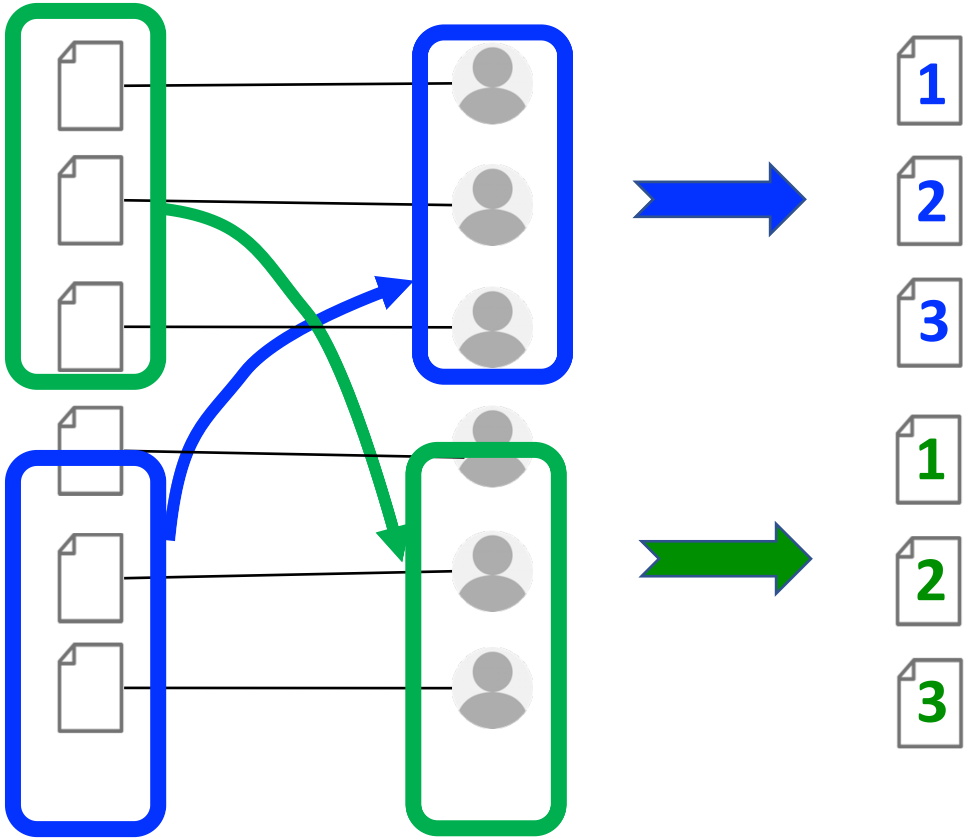
Given: Conflict graph
(e.g., authorship graph)



How to ensure that no reviewer can influence decision of any conflicted paper?

Partitioning method

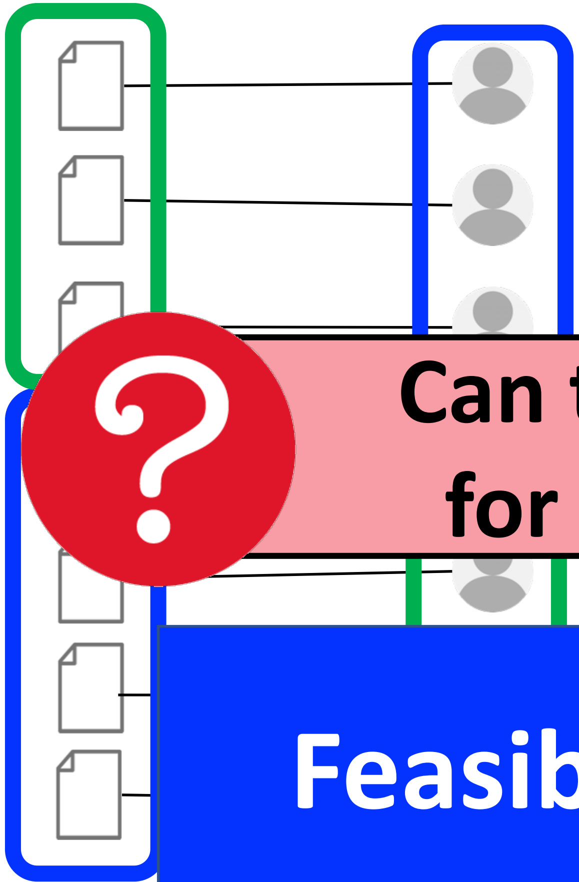
Primarily studied for peer grading



[Alon et al. 2011, Holzman et al. 2013, Bousquet et al. 2014, Fischer et al. 2015, Kurokawa et al. 2015, Kahng et al. 2017; see also Aziz et al. 2019, Mattei et al. 2020]

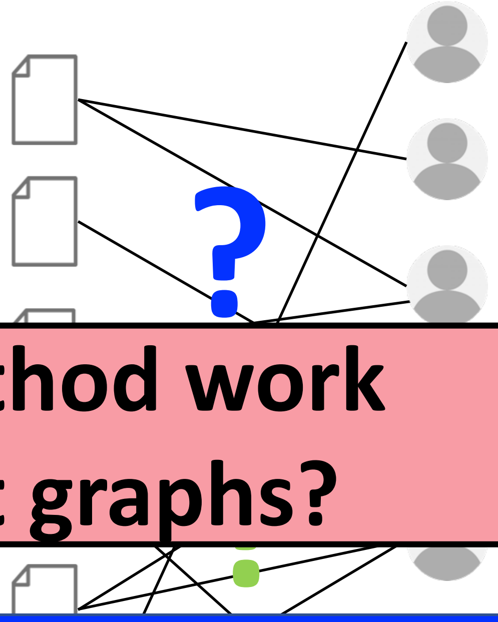
Peer grading

1-1 conflict graphs



Conference peer review

More **complex** conflict graphs



Can the partitioning method work for peer-review conflict graphs?

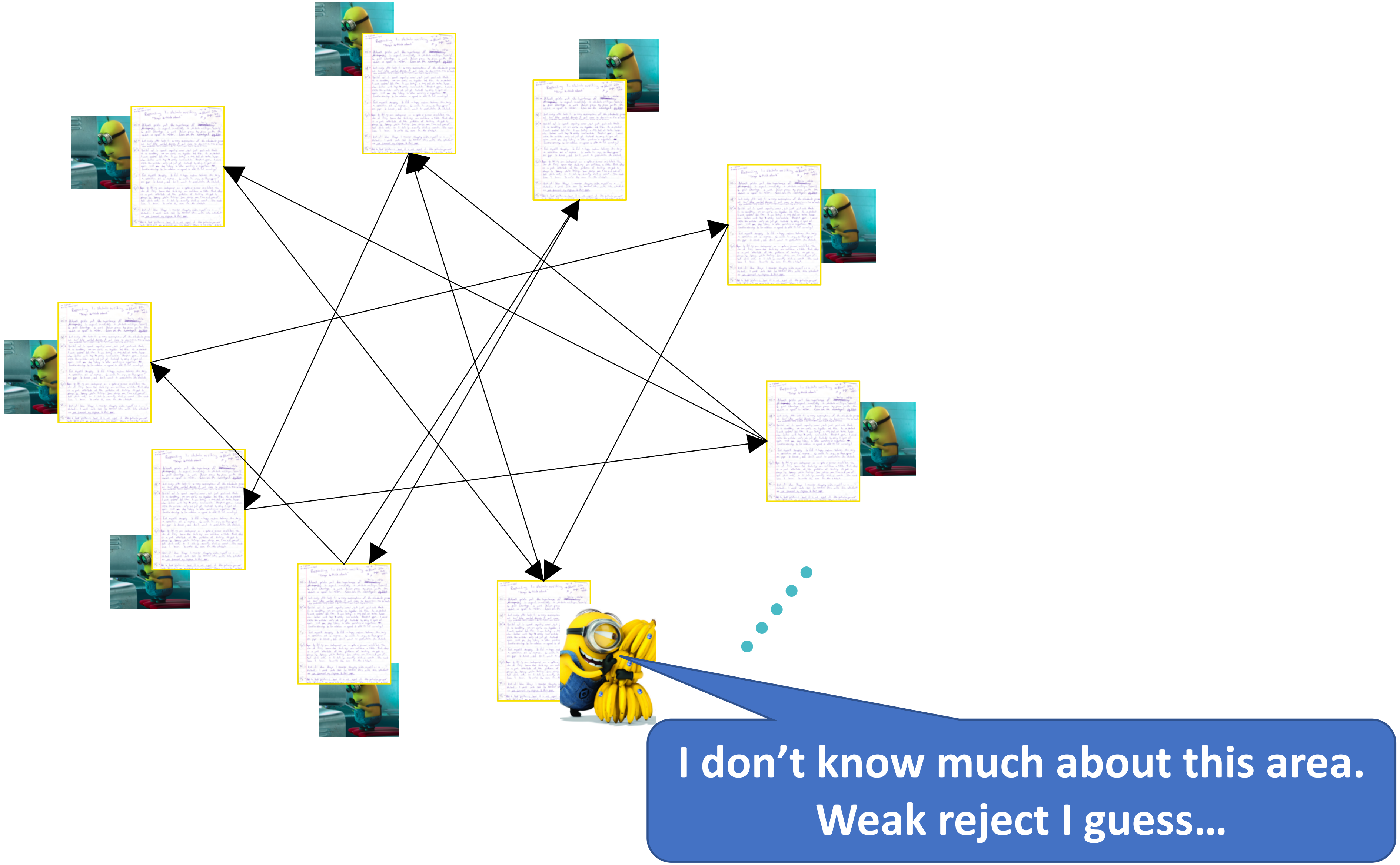
Feasible, but loss of efficiency.

[Xu et al. 2018]

Dishonest behavior: Open problems

- Detect/prevent other forms of dishonest behavior
  [Ferguson et al. 2014, Gao et al. 2017, Langford 2008]

Noise



Noise and reviewer assignment

Poor reviews due to inappropriate choice of reviewers

"one of the first and **potentially most important** stages is the one that attempts to distribute submitted manuscripts to competent referees." [Rodriguez et al. 2007]

Top reason for dissatisfaction: "Reviewers or panelists not expert in the field, poorly chosen, or poorly qualified" [McCullough 1989]

Automated assignment

(Used in AAAI, NeurIPS, ICML,…)

Compute similarities

[Mimno et al. 2007, Rodriguez et al. 2008, Charlin et al. 2013, Liu et al. 2014]

Assignment

- For every pair (paper p, reviewer r), similarity score $\boldsymbol{s_{pr}} \in \mathbf{[0, 1]}$
- Based on
  - Match text of submitted paper with reviewer's past papers
  - Match chosen subject areas
  - Bids
- Higher similarity score ⇒ Better envisaged quality of review

- Use similarity scores to assign reviewers to papers…

Assignment: Maximize total similarity

(Used in AAAI, NeurIPS, ICML,…)

$$\underset{\text{assignment}}{\text{maximize}} \sum_{p\,\in\,\text{Papers}} \; \sum_{r\,\in\text{Reviewers}} s_{pr}\, \mathbb{I}\{\text{paper } p \text{ assigned to reviewer } r\}$$

subject to

Every paper gets at least certain #reviewers

Every reviewer gets at most certain #papers

No paper is assigned to conflicted reviewer

[Conference management systems: TPMS (Charlin and Zemel 2013), EasyChair, HotCRP]

[Goldsmith et al. 2007, Tang et al. 2010, Charlin et al. 2012, Long et al. 2013]

Toy example

- One reviewer per paper
- One paper per reviewer

	Paper A	Paper B	Paper C
Reviewer 1	**1**	**0**	**0.5**
Reviewer 2	**0.7**	**1**	**0**
Reviewer 3	**0**	**0.7**	**0**

Assignment is unfair to paper C

There exists another more balanced assignment

[Stelmakh et al. 2018]

Common approach: Maximize total similarity

$$\underset{\text{assignment}}{\text{maximize}} \sum_{p \in \text{Papers}} \sum_{r \in \text{Reviewers}} s_{pr} \, \mathbb{I}\{\text{paper i assigned to reviewer j}\}$$

- **Unbalanced:** Can assign all relevant reviewers to some papers and all irrelevant reviewers to others [Stelmakh et al. 2018]
- **Can be particularly unfair** to interdisciplinary papers
- On CVPR 2017 data, assigns at least one paper **all reviewers with 0 similarity** (there are other assignments that do much better) [Kobren et al. 2019]

More balanced assignment

$$\underset{\text{assignment}}{\text{maximize}} \quad \underset{p \in \text{Papers}}{\text{minimum}} \sum_{r \in \text{Reviewers}} s_{pr} \, \mathbb{I}\{\text{paper i assigned to reviewer j}\}$$

subject to

Every paper gets at least certain #reviewers

Every reviewer gets at most certain #papers

No paper is assigned to conflicted reviewer

Fix assignment for the worst-off paper $\underset{p \in \text{Papers}}{\text{argmin}}$

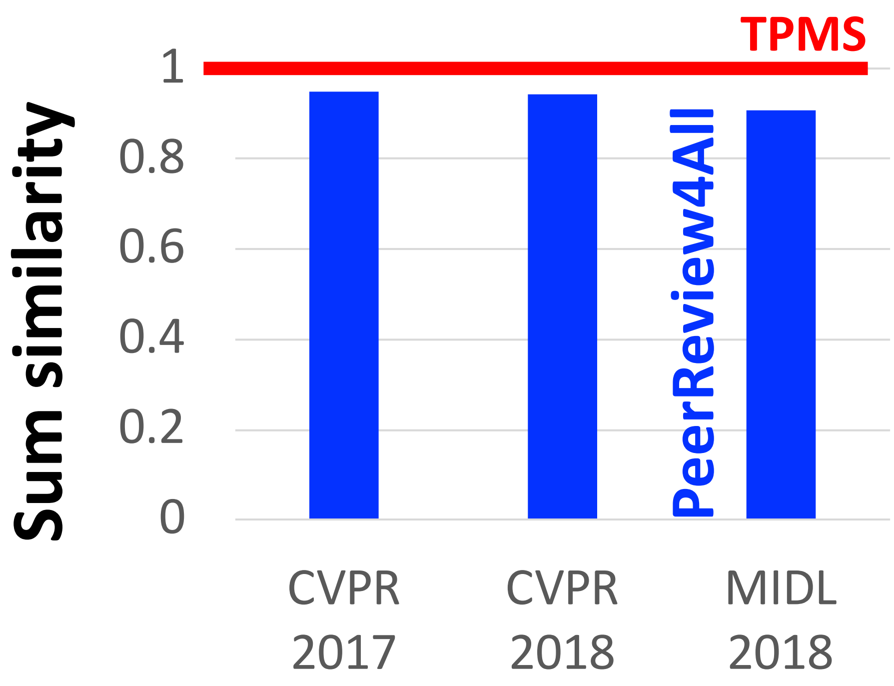
Repeat for remaining papers

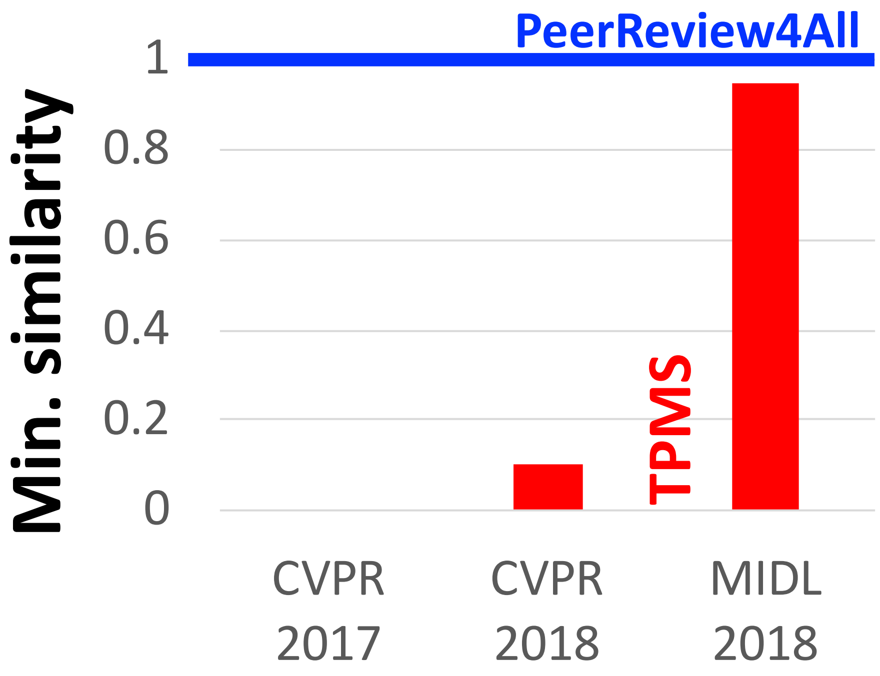
- NP Hard [Garg et al. 2010]
- Approximation algorithm ("PeerReview4All")
- Statistical guarantees on overall top-K selection

[Stelmakh et al. 2018]

Evaluation

- **TPMS algorithm** optimizes **sum similarity**
- **PeerReview4all algorithm** [Stelmakh et al. 2018] optimizes **minimum similarity**





[Evaluations by Kobren et al. 2019]

- PeerReview4All used in ICML 2020: Outcome similar to above

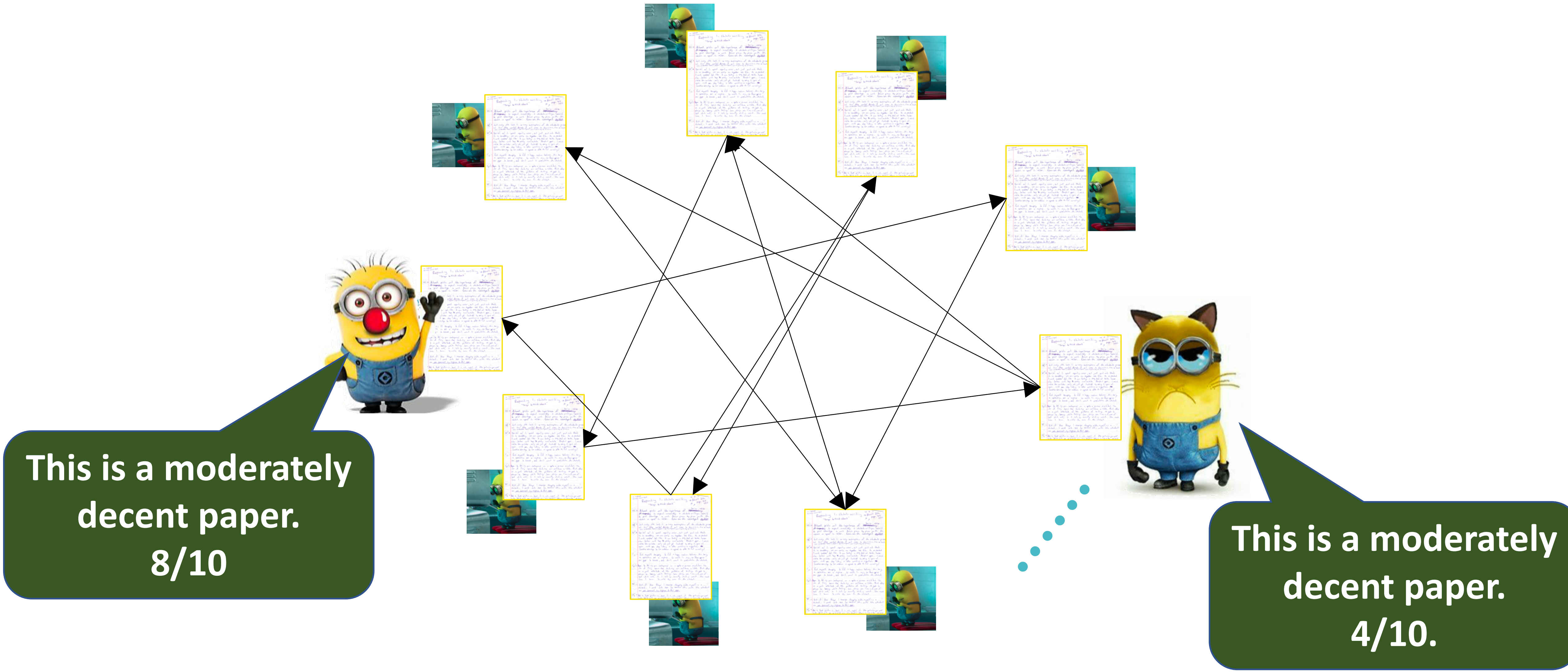
Noise: Open problems

- Better computation of similarities; joint similarity computation and assignment

[Mimno et al. 2007, Rodriguez et al. 2008, Charlin et al. 2013, Liu et al. 2014, Tran et al. 2017]

- Denoise using text of reviews

Miscalibration



Miscalibration in ratings

"A raw rating of 7 out of 10 in the absence of any other information is potentially useless." [Mitliagkas et al. 2011]

"The rating scale as well as the individual ratings are often arbitrary and may not be consistent from one user to another." [Ammar et al. 2012]

"[Using rankings instead of ratings] becomes very important when we combine the rankings of many viewers who often use completely different ranges of scores to express identical preferences." [Freund et al. 2003]

Unfairness in peer review

“the existence of disparate categories of reviewers creates the potential for unfair treatment of authors. Those whose papers are sent by chance to assassins/demoters are at an unfair disadvantage, while zealots/pushovers give authors an unfair advantage.”

[Siegelman 1991]

Two approaches in the literature

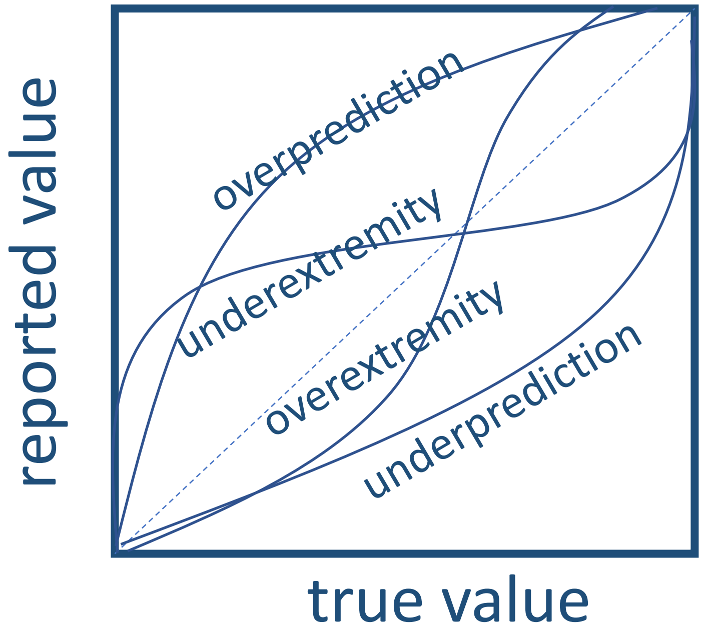
1 **Assume simplified (linear) models for calibration**

[Paul 1981, Flach et al. 2010, Roos et al. 2011, Baba et al. 2013, Ge et al. 2013, Mackay et al. 2017]

- Did not work well [NeurIPS 2016 program chairs; personal communication]
- *"We experimented with reviewer normalization and generally found it significantly harmful."* [Langford (ICML 2012 program co-chair)]

Miscalibration is quite complex:

[Brenner et al. 2005]



Two approaches in the literature

2 **Use rankings** [Rokeach 1968, Freund et al. 2003, Harzing et al. 2009, Mitliagkas et al. 2011, Ammar et al. 2012, Negahban et al. 2012]

- Use rankings induced by ratings or directly collect rankings
- Commonly believed to be the best option if no assumptions on miscalibration

? **Is it possible to do better using ratings than rankings, with essentially no assumptions on the miscalibration?**

Canonical 2x2 setting

$z_A^* \neq z_B^* \in [0,1]$

Miscalibration function: $f_1 \; : [0,1] \rightarrow [0,1]$

Given paper $i \in \{A, B\}$, outputs $f_1(z_i^*)$

Miscalibration function $f_2 \; : [0,1] \rightarrow [0,1]$

Given paper $i \in \{A, B\}$, outputs $f_2(z_i^*)$

- Adversary chooses z_A^*, z_B^* and strictly monotonic f_1, f_2
- **One paper assigned to each reviewer at random**
- **Goal: Given (assignment, score given by each reviewer)**

  estimate if $z_A^* > z_B^*$ or $z_B^* > z_A^*$

  - Eliciting rankings is vacuous; amounts to random guessing

[Wang et al. 2018]

Impossibility on deterministic estimators

Theorem

No deterministic estimator has a success probability better than ranking.

[Wang et al. 2018]

A randomized estimator

Theorem

There is a randomized estimator that strictly outperforms ranking.

[Wang et al. 2018]

With probability (1+|difference between the two scores|)/2, pick paper which received higher score

	Reviewer 1: $f_1(x) = x/2$	Reviewer 2: $f_2(x)=(3+x)/4$
Paper A: $z_A^* = 0.2$	$f_1(0.2) = 0.1$	$f_2(0.2) = 0.8$
Paper B: $z_B^* = 0.6$	$f_1(0.6) = 0.3$	$f_2(0.6) = 0.9$

- Under blue assignment, pick paper **B** with probability

$$\frac{1 + |0.1 - 0.9|}{2} = 0.9$$

(output is correct)

- Under red assignment, pick paper **A** with probability

$$\frac{1 + |0.3 - 0.8|}{2} = 0.75$$

(output is wrong)

- On average, correct with probability

$$\frac{1}{2}(0.9) + \frac{1}{2}(1 - 0.75) = 0.575 > 0.5$$

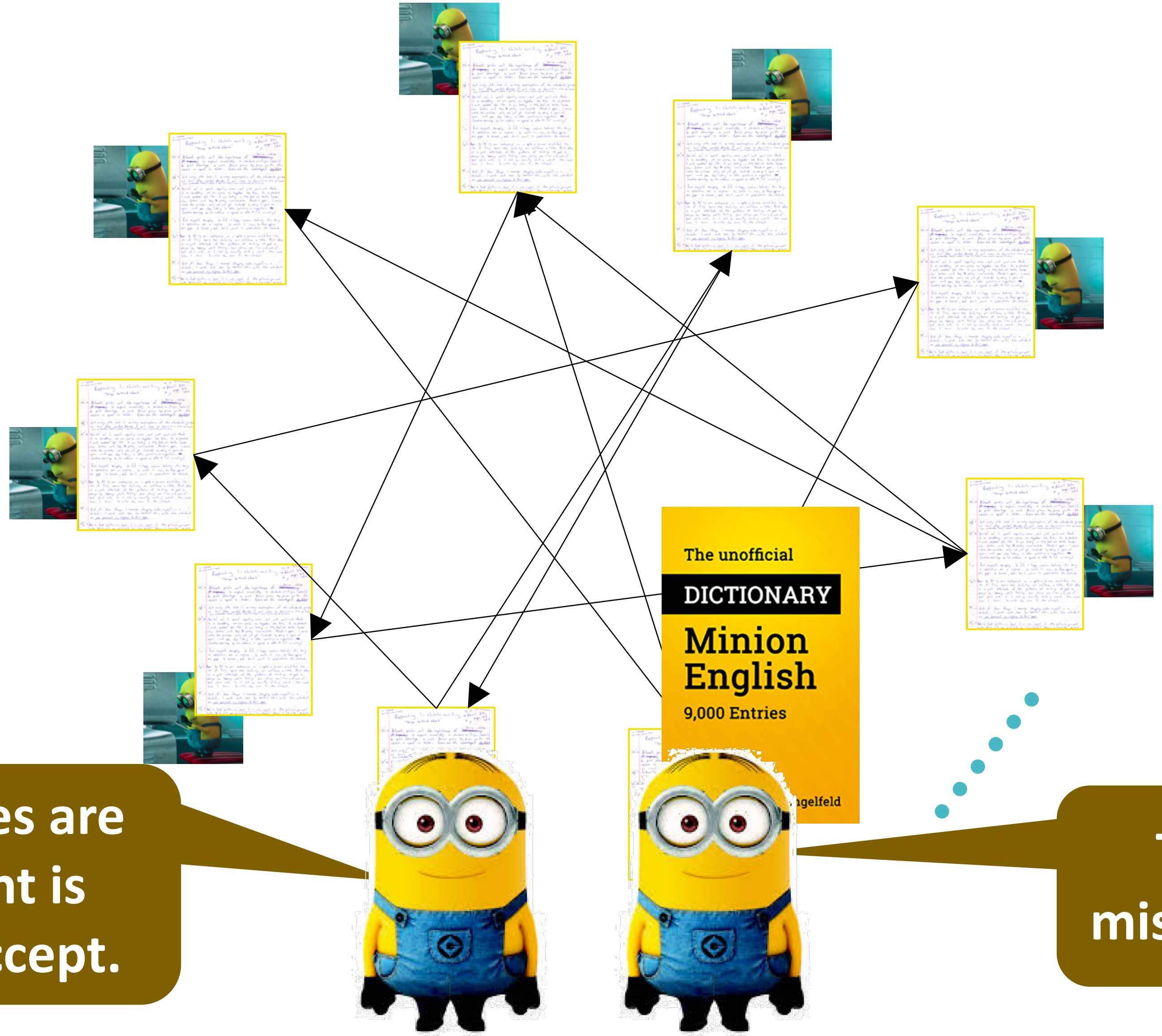
[Wang et al. 2018]

Miscalibration: Open problems

Strong assumptions: parametric, linear | **Sweet spot ?** | Arbitrary/adversarial miscalibration | Ranking

- Weaker assumptions: non-parametric, non linear (e.g., permutation-based models [Shah 2017 part 1])
- Amenable to small sample sizes: Avoid overkill

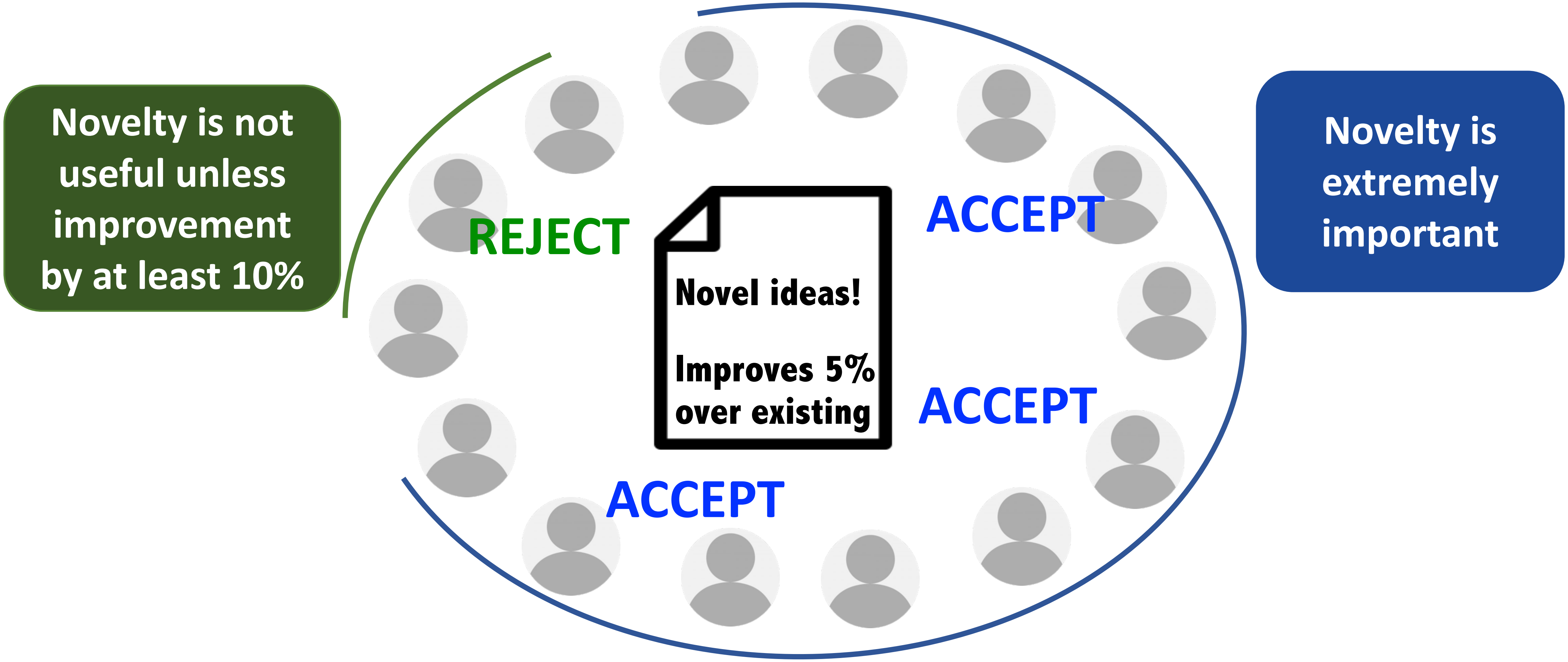
Subjectivity



Spelling mistakes are ok. The content is great. Strong accept.

Too many spelling mistakes. Strong reject.

Differing opinions about relative importance of criteria



[Kerr et al. 1977, Bakanic et al. 1987, Hojat et al. 2003, Church 2005, Lamont 2009, Lee 2015]

Commensuration Bias in Peer Review

Carole J. Lee*†

To arrive at their final evaluation of a manuscript or grant proposal, reviewers must convert a submission's strengths and weaknesses for heterogeneous peer review criteria into a single metric of quality or merit. I identify this process of commensuration as the

"Illuminates how intellectual priorities in individual peer review judgments can collectively subvert the attainment of community-wide goals"

[Lee 2015]

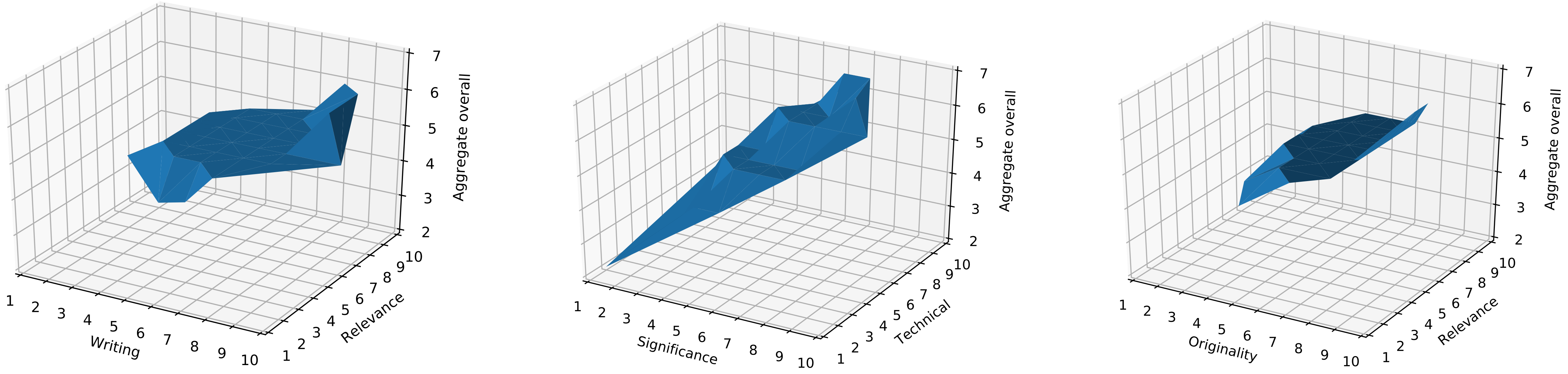
How to ensure that every paper is judged by the same yardstick?

Using ML and social choice theory

Learn a mapping from criteria to overall scores based on all reviews, and apply this mapping to all reviews.

[Noothigattu et al. 2018]

IJCAI 2017



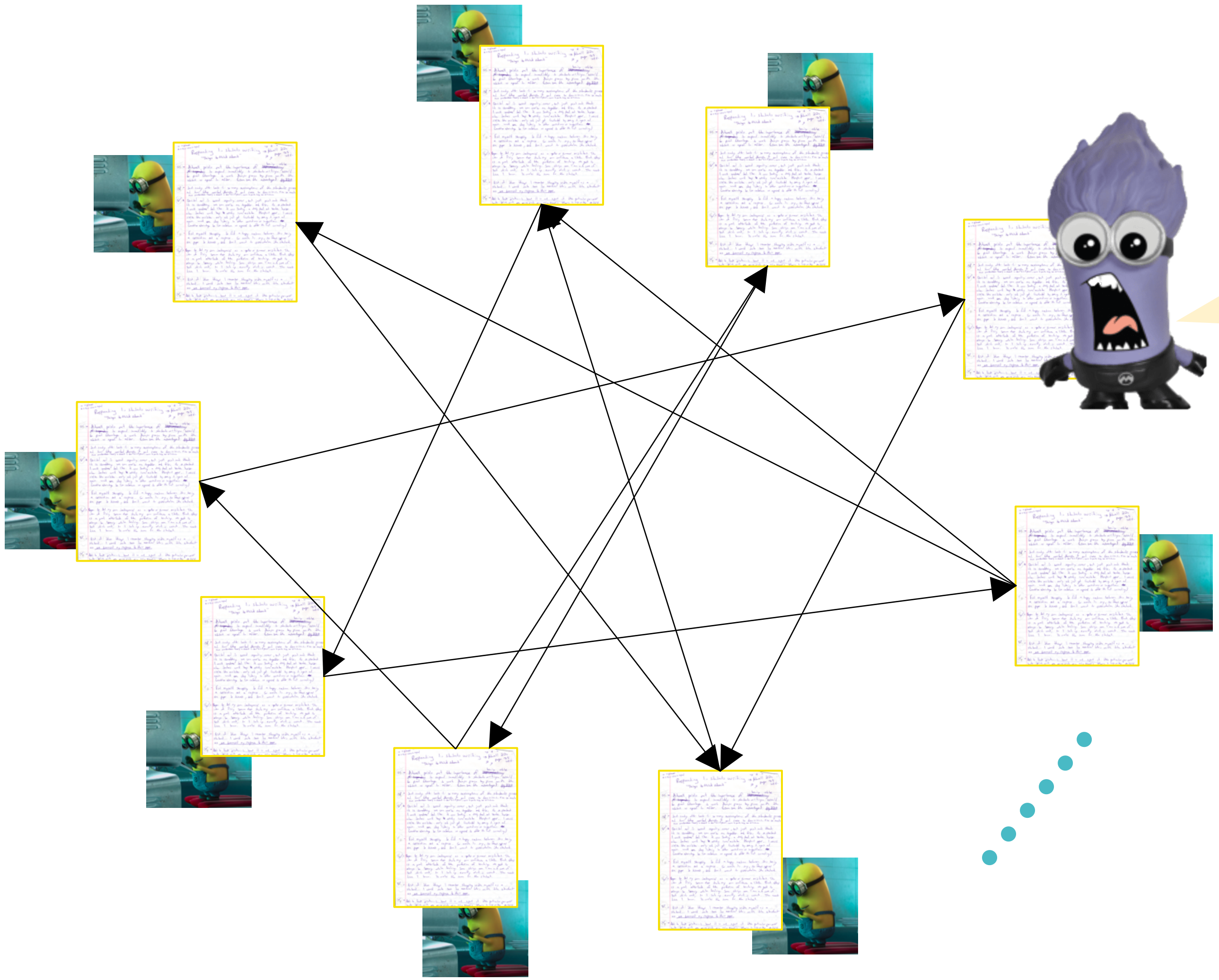
- Writing and Relevance: Really bad - significant downside, really good - appreciated, in between - irrelevant.
- Technical quality and Significance: high influence; the influence is approximately linear.
- Originality: moderate influence.

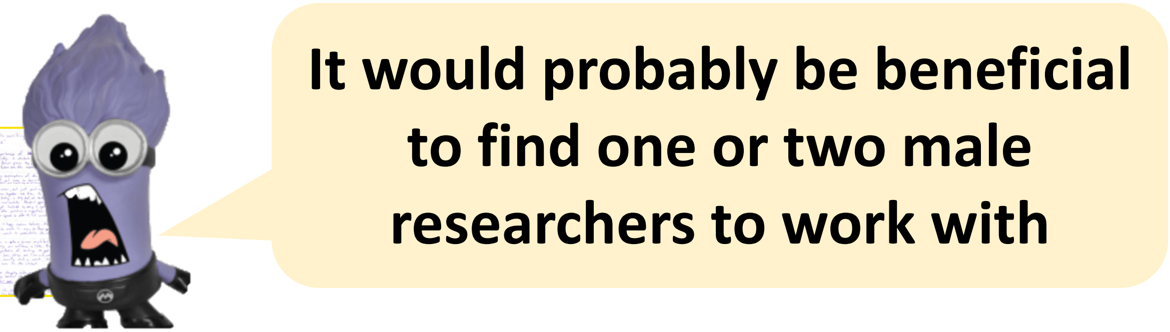
[Noothigattu et al. 2018]

Subjectivity: Open problems

- Evaluation in absence of ground truth

Biases





True story

Review in PLOS ONE, 2015

Authors: Fiona Ingleby, Megan Head

Single blind versus double blind

A Principled Interpretation of Minion Speak

S. Overkill and F. Gru
Cartoony Minion University

In this paper we present a new understanding of…

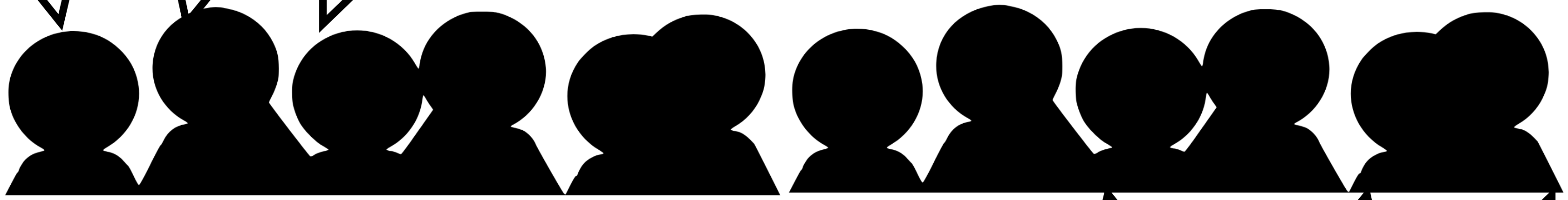
A Principled Interpretation of Minion Speak

Anonymous Authors
Anonymous Affiliation

In this paper we present a new understanding of…

Lot of debate!

Single blind can lead to gender/fame/race/… biases

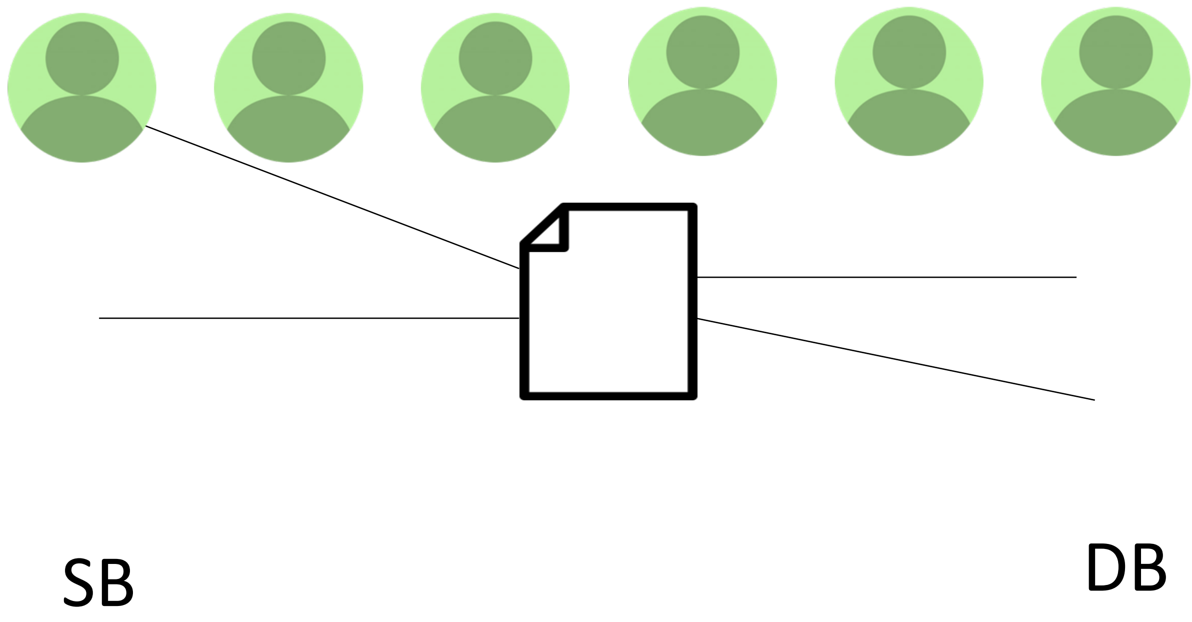


Where is the evidence of bias in my research community?

How to rigorously test for biases in peer review (while ensuring “good” review process)?

WSDM'17 experiment: Setup

A remarkable experiment!



- Reviewers randomly split into single blind (SB) and double blind (DB) conditions
- Each paper assigned 2 SB reviewers and 2 DB reviewers

[Tomkins et al. 2018]

WSDM’17 experiment: Tests for bias regarding…

- Gender
- Famous author
- Top university
- Top company
- From USA
- Academic institution
- Reviewer same country as author

[Tomkins et al. 2018]

WSDM'17 experiment: Findings

- Famous author
- Top university
- Top company

Significant bias

- At least one woman author

Not statistically significant; high effect size
Meta analysis is statistically significant

- From USA
- Academic institution
- Reviewer same country as author

No evidence of bias

WSDM moved to double blind from the following year.

[Tomkins et al. 2018]

Peculiar characteristics of peer review

[Stelmakh et al., 2019]

Statistical testing preliminaries

False alarm (Type I error) Claiming **presence** of bias when the bias is **absent**

Detection (1 - Type-II error) Claiming **presence** of bias when the bias is **present**

For a given α, must ensure
P(false alarm) $\leq \alpha$

Typical choice: α = 0.05

[Stelmakh et al., 2019]

Characteristic 0: Correlations between quality of papers and certain attributes

- Famous author
- Top university
- Top company

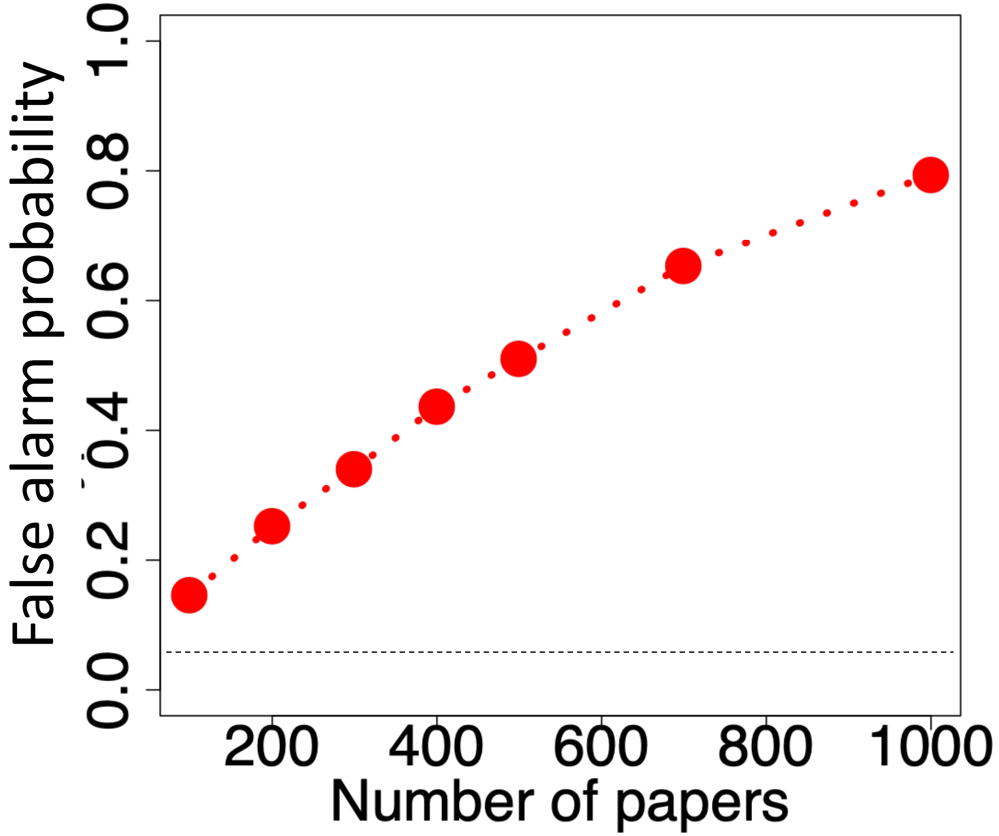
Combined with other characteristics…

[Stelmakh et al., 2019]

Characteristic 1: Reviews are noisy

Reviewers are imperfect (noisy)

Must ensure: P(declare bias when no bias) $\leq$ 0.05

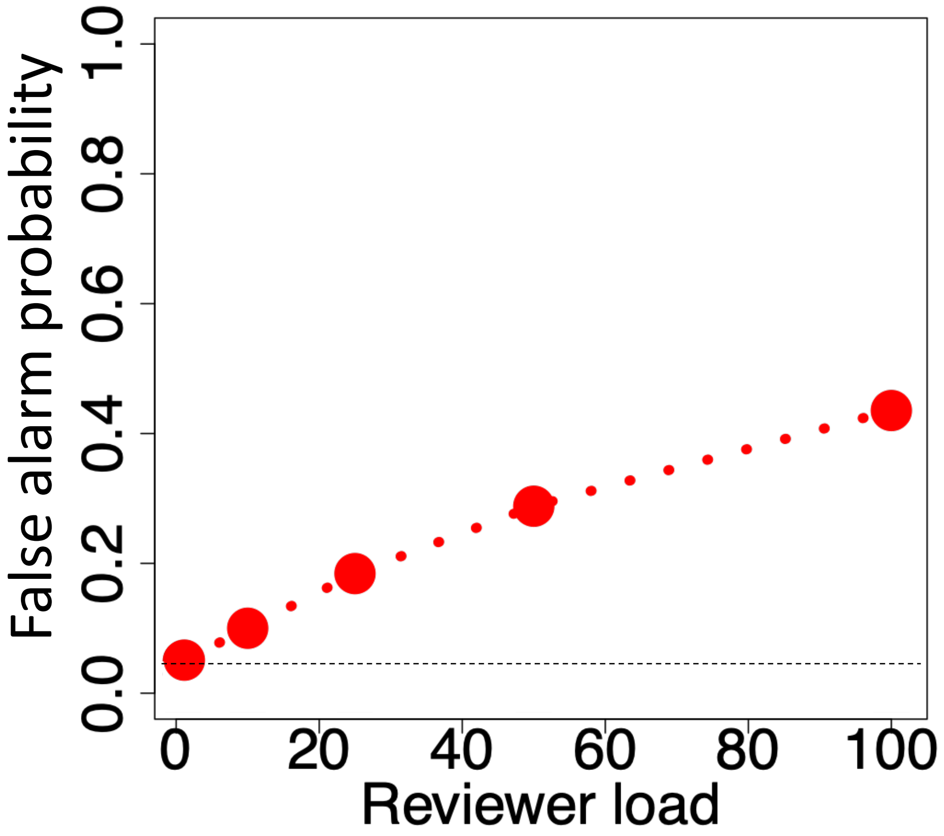


[Stelmakh et al., 2019]

Characteristic 2: Intra-reviewer dependency

Reviews of different papers by the same reviewer are dependent, e.g., a reviewer may be lenient or strict

Must ensure: P(declare bias when no bias) ≤ 0.05

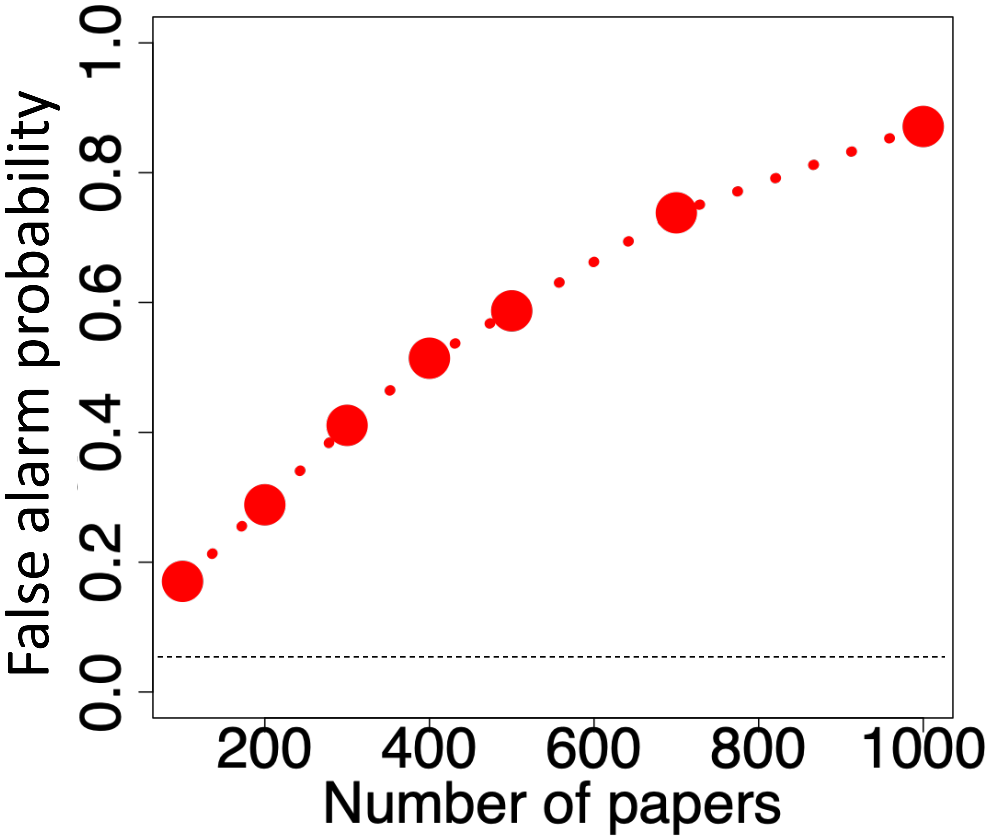


[Stelmakh et al., 2019]

Characteristic 3: Model complexity

Human evaluations may be more complex than simple parametric/logistic models

Must ensure: P(declare bias when no bias) $\leq$ 0.05

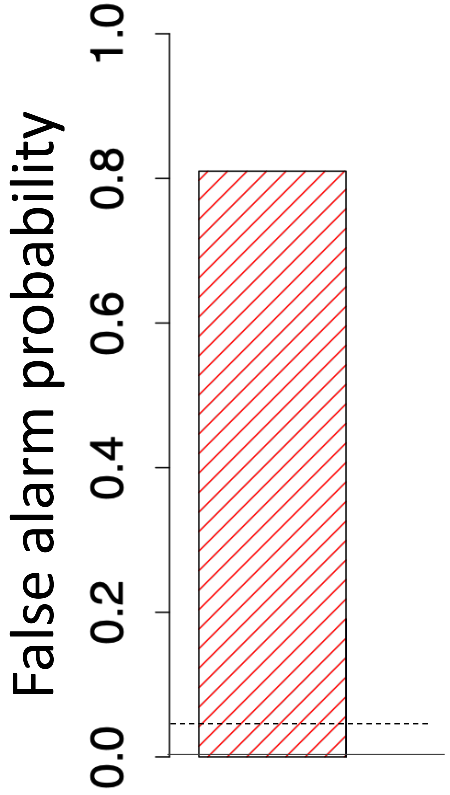


[Stelmakh et al., 2019]

Characteristic 4: Non-random assignment

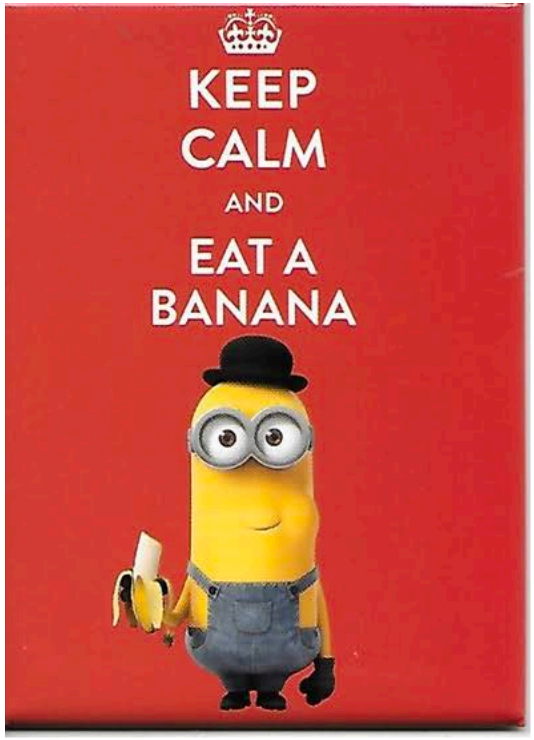
Assignment of reviewers to papers is NOT random

Must ensure: P(declare bias when no bias) ≤ 0.05



[Stelmakh et al., 2019]

These issues are fixable!



- General framework
- Careful modification of experimental setup
- Non-parametric test
- Strong theoretical guarantees:
  - False alarm control
  - Non-trivial detection power

[Stelmakh et al., 2019]

Biases: Open problems

- Tests of biases from observational peer-review data
- Biases in other review components such as program committee meetings and discussions

Norms and Policies

Alright, so here's what everyone must do…

Biases due to alphabetical ordering

In Economics, norm is to order authors in alphabetical order of last names.

Faculty with last name starting with an earlier alphabet are:

- Significantly more likely to receive tenure
- Significantly more likely to become fellows of the Econometric Society
- More likely to receive the Clark Medal and the Nobel Prize

The (related) field of Psychology, which does not order by alphabet, does not show any of these biases.

[Einav and Yariv 2006] (See also [Hilmer et al. 2005, van Praag et al. 2008, Ray et al. 2018])

What causes these biases?

In papers

Implicit bias – Primacy effects

Explicit bias – “*First author* et al.”

Conference	#Total papers	#Papers using “*First author* et al.” in its text
STOC 2017	99	70
STOC 2016	79	59
FOCS 2017	79	48
FOCS 2016	73	43
EC 2017	75	48
EC 2016	99	87

On websites

Serial position effects

PC Members

Aaron J Elmore (University of Chicago)
Abdussalam Alawini (University of Illinois at U
Alan Fekete (University of Sydney)
Alex Beutel (Google)
Alexander Boehm (SAP SE)
Alexandra Meliou (University of Massachuset
Alexandros Labrinidis (University of Pittsburg
Alin Deutsch (UCSD)
Alvin Cheung (UC Berkeley)

https://researchonresearch.blog/2018/11/28/theres-lots-in-a-name/

Let's fix this!

In papers

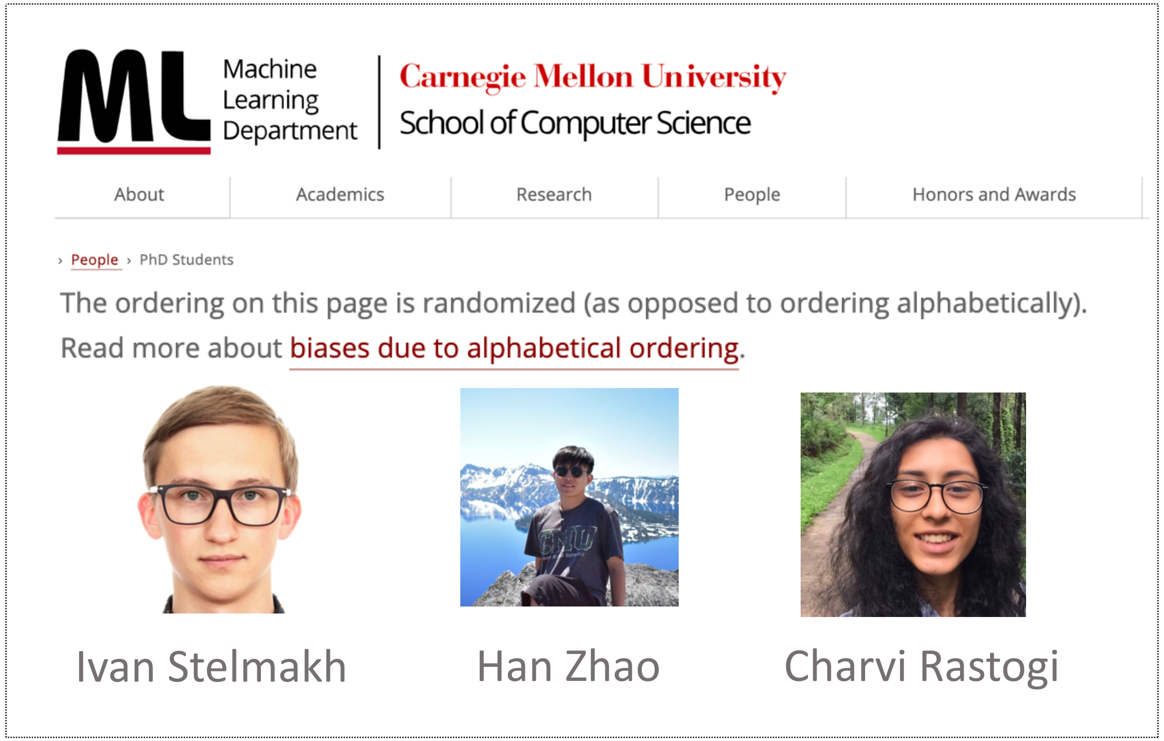
ACM EC conference now uses numbering instead of “first author et al.” citation style

Can randomize author ordering

On websites

CMU Machine Learning Department website now uses dynamic randomization for ordering people

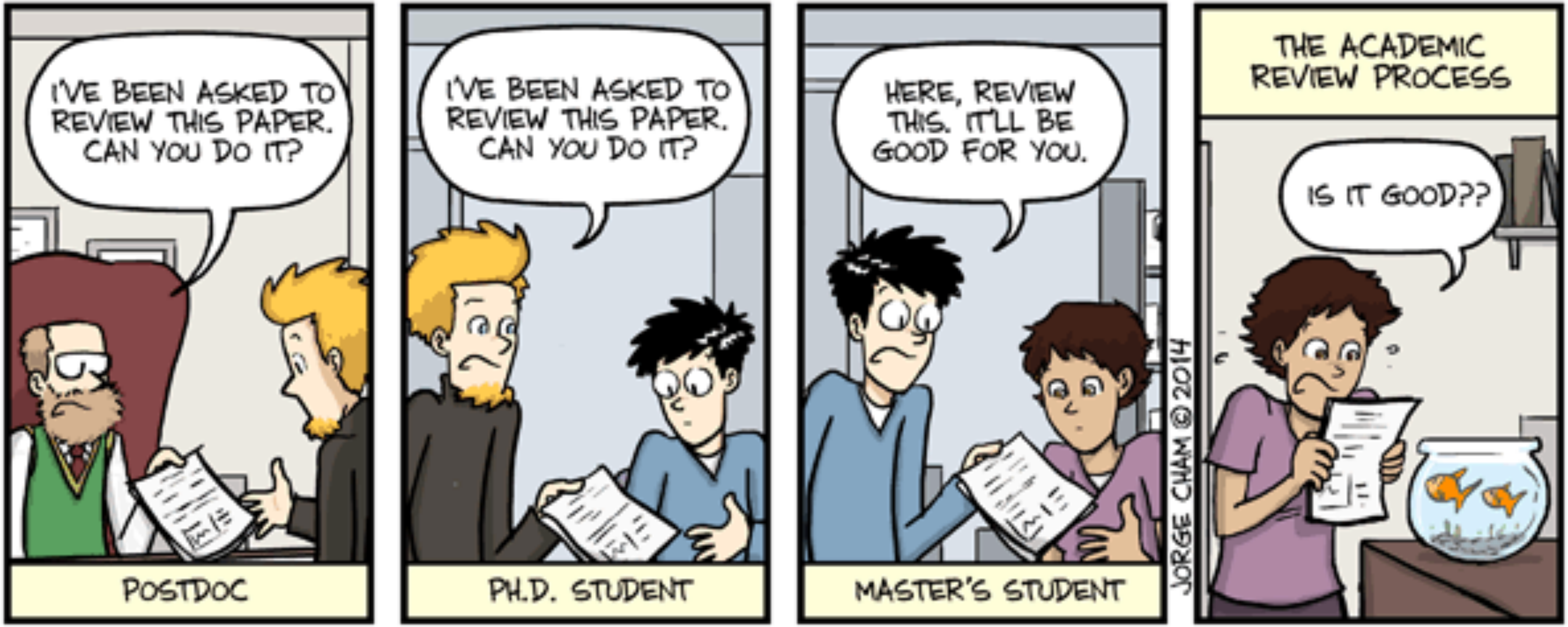
www.ml.cmu.edu/people/phd-students.html



https://researchonresearch.blog/2018/11/28/theres-lots-in-a-name/

Conclusions

- **Many sources of biases and unfairness in peer review**
- **Urgent need to revamp peer review, at scale**
  - Lot at stake: Careers, Scientific progress
- **Lots of open problems!**
  - Exciting
  - Theoretical / Applied / Conceptual
  - Challenging
  - **Impactful**



"Piled Higher and Deeper" by Jorge Cham

Thank you! Questions?

http://cs.cmu.edu/~nihars
nihars@cs.cmu.edu

“**Today reviewing is like grading**: When grading exams, zero credit goes for thinking of the question. When grading exams, zero credit goes for a novel approach to solution. (Good) reviewing: acknowledges that the question can be the major contribution. (Good) reviewing: acknowledges that a novel approach can be more important than the existence of the solution.” [Naughton 2010]

Backup slides: Biases

A more general formulation

Absence of bias:

P(reviewer *r* accepts paper *p* in **SB**) **=** P(reviewer *r* accepts paper *p* in **DB**)

Presence of bias:

$$\text{P(reviewer } r \text{ accepts paper } p \text{ in } \mathbf{SB}) \underset{p \,\notin\, \text{Group}}{\overset{p \,\in\, \text{Group}}{\lesseqgtr}} \text{P(reviewer } r \text{ accepts paper } p \text{ in } \mathbf{DB})$$

and at least one inequality is strict.

- No assumption of existence of any “true scores”
- Non-parametric model
- More general view greatly simplifies things…

[Stelmakh et al., 2019]

Experimental setup and test

Step 1: Experimental setup (Reviewer assignment)

(1a) Initial assignment: Each paper assigned 2 reviewers; at most 1 paper per reviewer

(1b) Randomization: For each paper, send 1 reviewer to SB and 1 to DB uniformly at random

(1c) Final assignment: Assigning remaining reviewers in any manner desired

Step 2: Statistical test (after getting reviews)

- Condition on triples from (1a) where reviewers disagree on their decisions
- Run permutation test at the level α

[Stelmakh et al., 2019]

Guarantees

Theorem (informal)

This experimental setup and test controls the false alarm probability at any given level $\alpha \in (0,1)$ and has asymptotic probability of detection of 1.

[Stelmakh et al., 2019]

Backup slides: Noise

Another example

- Two reviewers per paper
- One paper per reviewer

	Paper A	Paper B	Paper C
Reviewer 1	**0.9**	**0**	**0.5**
Reviewer 2	**0.6**	**0**	**0.5**
Reviewer 3	**0**	**0.9**	**0.5**
Reviewer 4	**0**	**0.6**	**0.5**
Reviewer 5	**0**	**0**	**0**
Reviewer 6	**0**	**0**	**0**
Total:	**1.5**	**1.5**	**0.0**
Total:	**0.9**	**0.9**	**1.0**

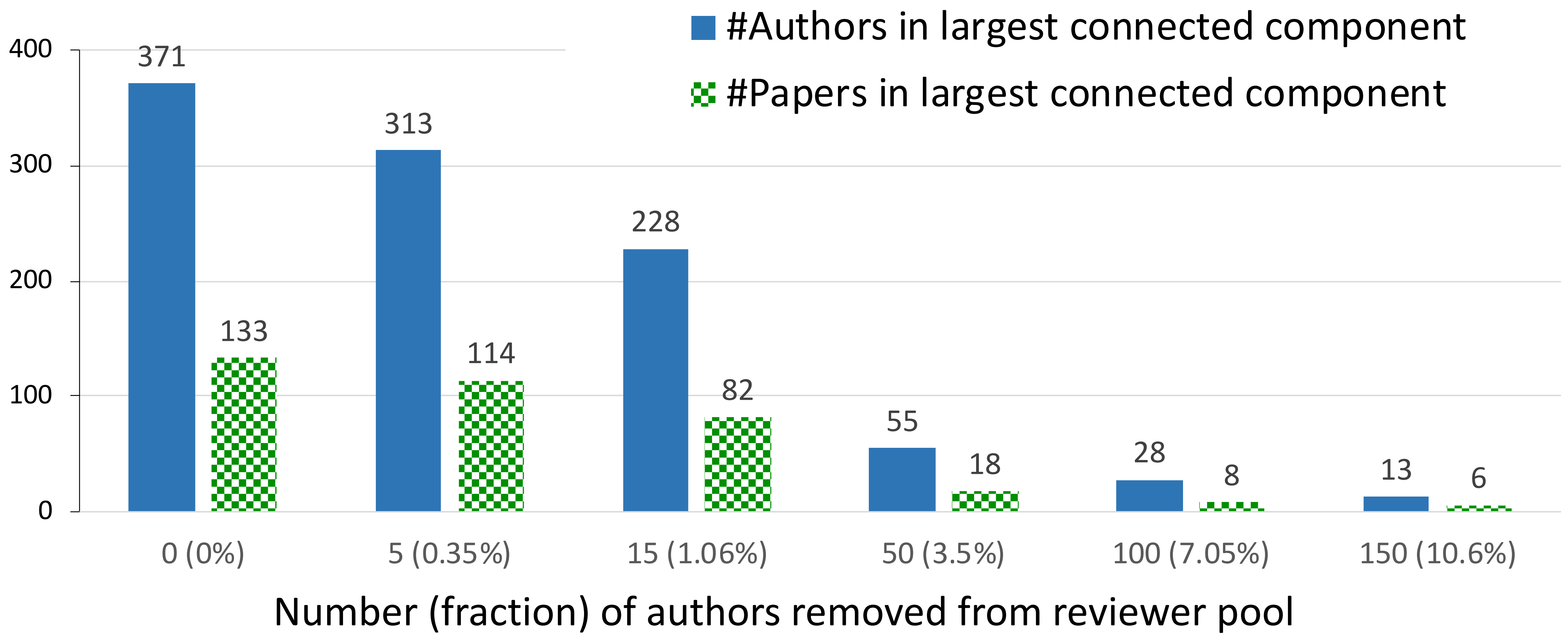
Assignment is unfair to (inter-disciplinary) paper C

There exists another more balanced assignment

[Stelmakh et al. 2018]

Backup slides: Dishonest behavior

Heuristic: Remove few authors from reviewer pool



[Xu et al. 2018]

Backup slides: Miscalibration

NeurIPS 2016

	1 (low or very low)	2 (sub-standard)	3 (poster level: top 30%)	4 (oral level: top 3%)	5 (award level: top 0.1%)
Impact	6.5%	36.1%	45.7%	10.5%	1.1%
Quality	6.7%	38.0%	44.7%	9.5%	1.1%
Novelty	6.4%	34.8%	48.1%	9.7%	1.1%
Clarity	7.1%	28.0%	48.6%	14.6%	1.8%

≥ 3: **57%** instead of intended 30%
≥ 4: **10%** instead of intended 3%
≥ 5: **1%** instead of intended 0.1%

[Shah et al. 2017]

Backup slides: Subjectivity

An axiomatic approach

- Challenge: There is no ground truth!

- Axiomatic approach
  - Approach is popular in economics and social choice theory
  - Identify "special case" scenarios that is easy to reason about
  - Establish necessary conditions (or "axioms")

[Noothigattu et al. 2018]

Special case

Notation: Reviewer i gives to paper j

- Criteria scores $x_{ij} \in [0,1]^k$
- Overall score $y_{ij} \in [0,1]$

Consider the L(p,q) loss family (which is a popular, natural matrix-extension of the L(p) loss family for vectors). Here $p \in [1, \infty]$, $q \in [1, \infty]$

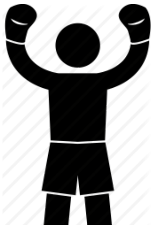
Special case: All reviewers review all papers. Moreover, any paper gets the same criteria scores from all reviewers.

Paper 1: $x_{11} = x_{21} = x_{31} = \cdots := x_1$
Paper 2: $x_{12} = x_{22} = x_{32} = \cdots := x_2$
$\vdots$

[Noothigattu et al. 2018]

Axiom 1: Consensus

For some $x \in [0,1]^k$ and $y \in [0,1]$, if all reviewers map x to y then $\hat{f}(x) = y$.



Axiom 2: Dominance

[Informal] If a paper a is *"at least as good as"* paper b, then $\hat{f}(x_a) \geq \hat{f}(x_b)$.

Axiom 3: Strategyproofness

No reviewer can bring the learnt overall scores closer to her/his own opinion by strategic manipulation. For any reviewer i, let $(y_{i1}, \ldots, y_{im})$ be overall scores she/he gives if honest. Let $\hat{f}$ denote learnt mapping in that case. Let $(y'_1, \ldots, y'_m)$ be any other overall scores and $\hat{g}$ be the associated learnt mapping. Then we need:

$$\left\| \left(\hat{f}(x_1), \ldots, \hat{f}(x_m) \right) - (y_{i1}, \ldots, y_{im}) \right\| \leq \left\| \left(\hat{g}(x_1), \ldots, \hat{g}(x_m) \right) - (y_{i1}, \ldots, y_{im}) \right\|$$

[Noothigattu et al. 2018]

Theorem

L(1,1) is the only L(p,q) loss that satisfies the three axioms.

- Strategyproofness violated when $q \in (1, \infty]$
- Consensus violated when $p = \infty$ and $q = 1$
- Dominance violated when $p \in (1, \infty)$ and $q = 1$

[Noothigattu et al. 2018]

Dominance violated under L(2,1) loss

- 2 papers, 3 reviewers, k=2 criteria
- Criteria scores x_1 = [¼, ¾], x_2 = [¾, ¼]
- Overall scores:

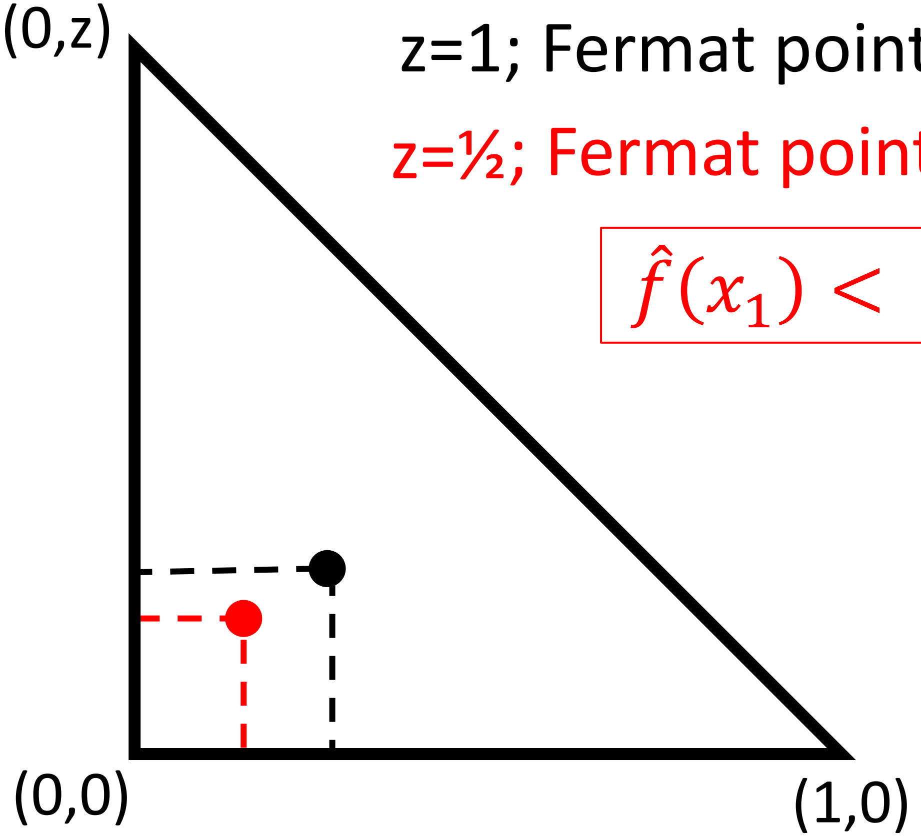
	Paper 1	Paper 2
Rev. 1	0	0
Rev. 2	1	0
Rev. 3	0	z < 1

Paper 1 dominates paper 2

$\boxed{\text{want } \hat{f}(x_1) \geq \hat{f}(x_2)}$

Fermat point of a triangle: Point with smallest total Euclidean distance from the 3 vertices

$(\hat{f}(x_1), \hat{f}(x_2))$ is exactly the Fermat point of:



z=1; Fermat point: (.20,.20)

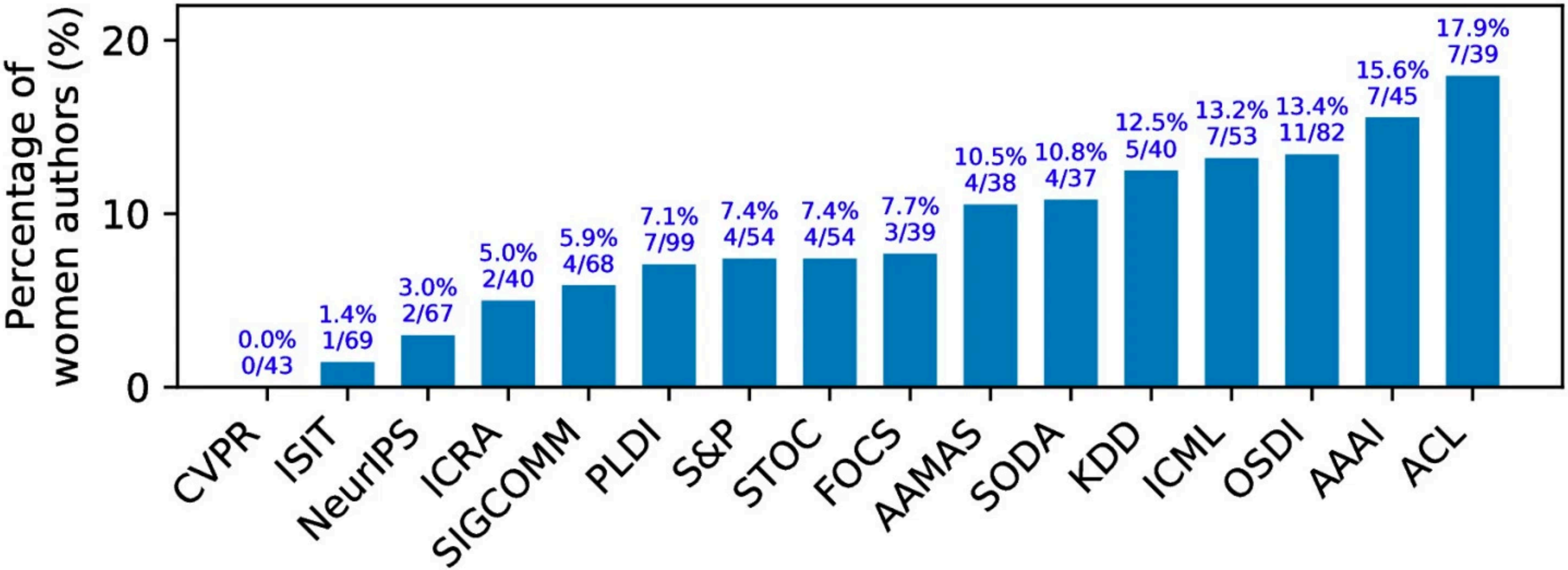
z=½; Fermat point: (.12,.15)

$\boxed{\hat{f}(x_1) < \hat{f}(x_2)}$

[Noothigattu et al. 2018]

Gender distribution in paper awards

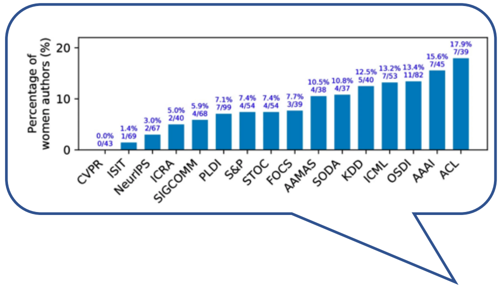
(2010– 2018)



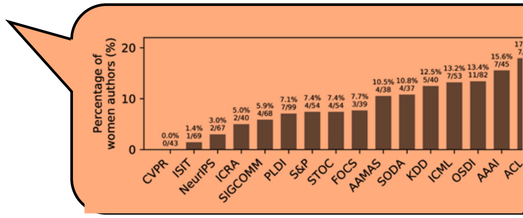
https://researchonresearch.blog/2019/06/18/gender-distributions-of-paper-awards/

Need for transparency

- Are author identities visible to the award committee?
- How is the committee determined?
- What criteria are used?



Started conversations in information theory society, NLP community, ML community, vision community,…



https://researchonresearch.blog/2019/06/18/gender-distributions-of-paper-awards/