POPL 2012 PC Chair Remarks

Michael Hicks

University of Maryland,

College Park, USA

Changes to POPL reviewing this year

- *Light* double blind reviewing (DBR)
- Extended Review Committee (ERC)
- Guardians

- Staged availability of supplemental material
- Added review form metadata
- Two-dimensional bidding

Double blind reviewing

- Scientific review may have unfair bias
  - "Knee jerk" reaction to gender, affiliation, location, and other non-technical metadata may affect reviewing
- Double blind reviewing (DBR) aims to avoid this bias by hiding authorship
- But DBR has costs
  - can hurt the submitted paper, inhibit the post-submission scientific process, and make it more difficult to find expert reviewers
- Is there a middle ground?

Light DBR

- Intention: unbiased “first look”
- Minimal changes to manuscript
  - Redact author names from front page
  - Cite papers in third person
  - Specified not to:
    - Anonymize self citations
    - Change system names
- Minimal limitations on post-submission dissemination
  - Giving talks, posting on web page not forbidden
- Still need solution to finding experts

Extended Review Committee

- Used at PLDI, ASPLOS, ISMM
- Two main purposes
  - Review PC submissions
  - Have large body of experts committed to performing a small number of reviews
    - Bid along with PC, can take COIs into account
    - Multiple reviews means more context
    - Less running around at the last minute to find experts
- 60 member ERC, 26 member PC

Guardians

- Sometimes PC+ERC still does not have sufficient expertise (e.g., due to COIs)
  - And reviewers cannot solicit outside experts because blinded authorship hides potential conflicts
- Solution
  - Reveal authorship after submitting review
    - At which point bias is much diminished
    - Now may solicit outside reviews once authorship is known
  - **Guardian** for each paper
    - PC member for non PC papers; ERC member otherwise
    - Submits review by halfway into review process
    - Finds outside reviewer if internal expert unlikely

Two dimensional bidding

- PC Chair Goal: each paper should be reviewed by experts, but also interested outsiders
  - Papers that are accessible have broader impact
- Problem: one-dimensional bids about interest, not expertise (correlated, but not equal)
  - Don't want to assign only interested non-experts to a paper; would prefer at least one uninterested expert
- Solution: Have **Interest** score and **Expertise** score
  - Assignment attempts to maximize interest while ensuring at least one expert
  - Novel Min-Cost, Max-Flow algorithm

Process Timeline

- **Jul 8**: Authors submit (and Chair vets) title, COIs
- **Jul 11**: 205 full papers in, bidding begins
- **Jul 21**: Chair distributes assignment
  - 20-22 papers per PC member, 3-6 per ERC member
- **Aug 22**: Guardian deadline
- **Sep 14**: Reviews released, authors respond
  - 4 papers withdrawn at this stage
- **Sep 18**: On-line discussion
- **Sep 30, Oct 1**: PC meeting
  - ERC decides PC papers via on-line discussion

The PC at the meeting

The work and results

- 852 reviews for 205 papers
  - 4.2 reviews per paper, on average
    - minimum = 3, maximum = 6
  - PC members: 20-23 reviews
  - ERC members: 2-6 reviews
  - Solicited 50 outside experts (1-2 reviews each)
- 45 papers accepted
  - 3 PC papers (out of 17)

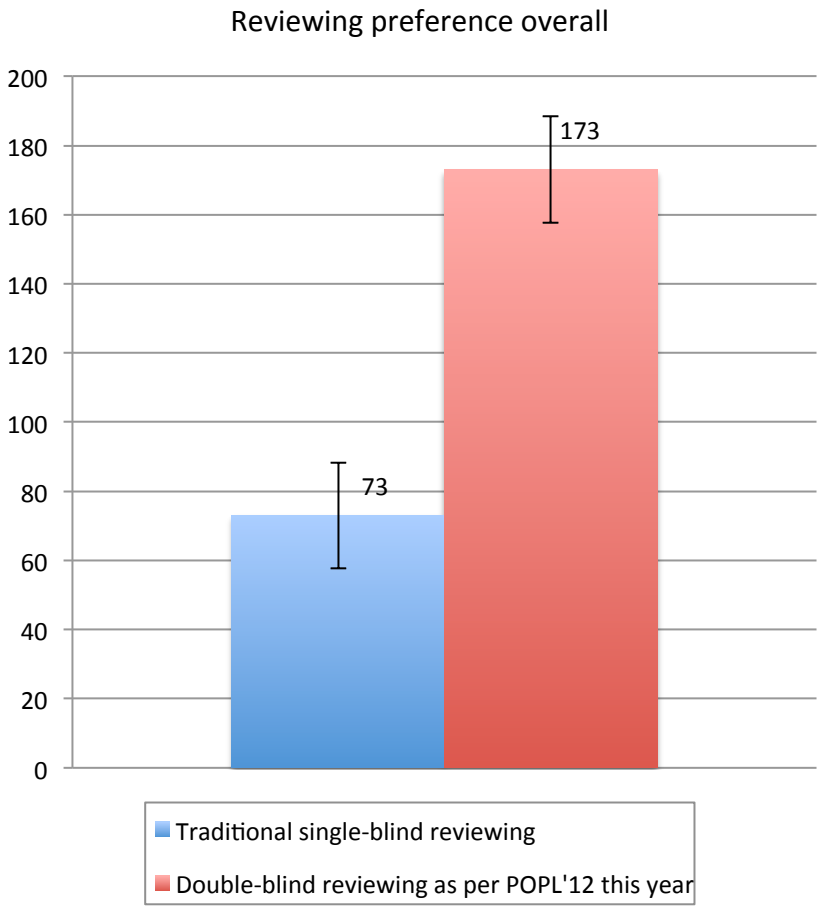
Was the process successful?

- Peer review is at the center of the scientific process. It is
  - a gateway for new ideas
  - the foundation of our trust in published results
- We want the highest quality, which means we need an effective, and fair, review process
- Ideally: we could measure outcomes directly
- In the meantime: survey participants

Three surveys

- PC/ERC pre-review survey
  - Opinion on DBR, various aspects of process
- PC/ERC post-decision survey
  - Opinion on DBR, elements of review process, efficacy of hiding authors
- Authors post-decision survey
  - Opinion and impact of DBR and other elements of the process
- Thanks to *Khoo* Yit Phang for the following charts

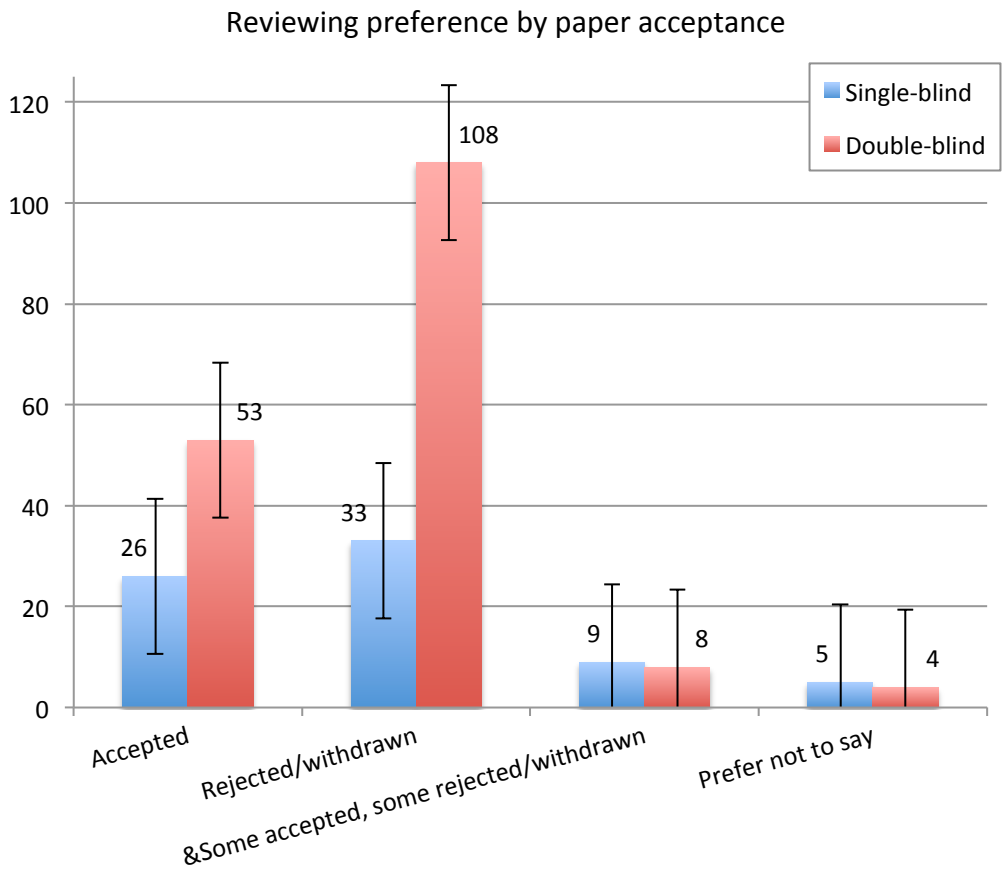
SBR or DBR: Authors



70% in favor of light DBR

30% in favor of traditional SBR

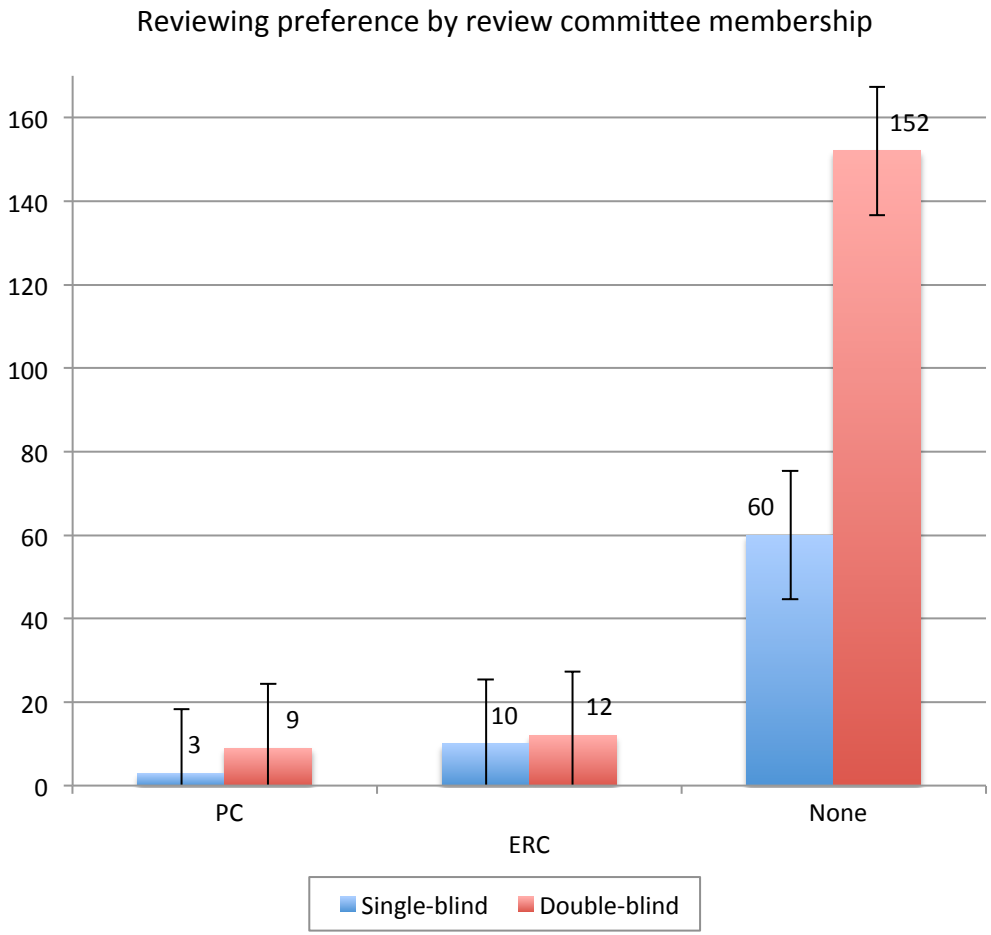
Correlating by paper outcome



(All) accepted papers: 67% +

(All) papers rejected: 77% +

Authors vs committee

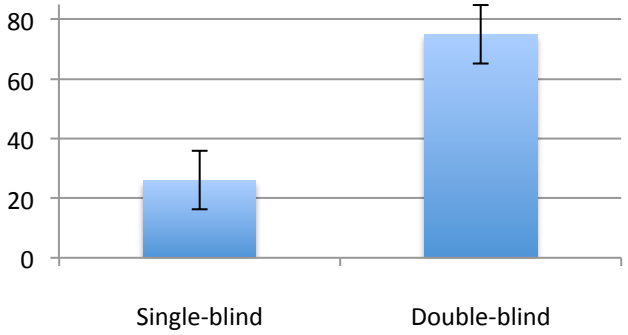


Non-committee authors: 72% +

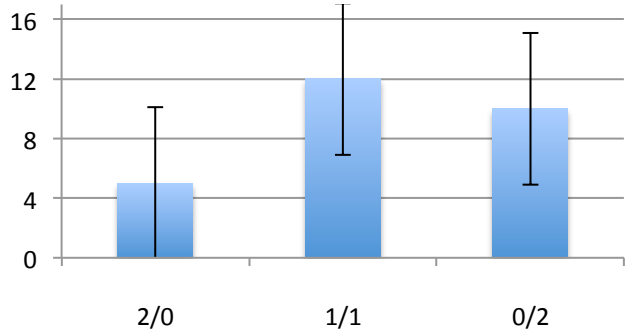
Separate survey for committee (next)

Authors disagree

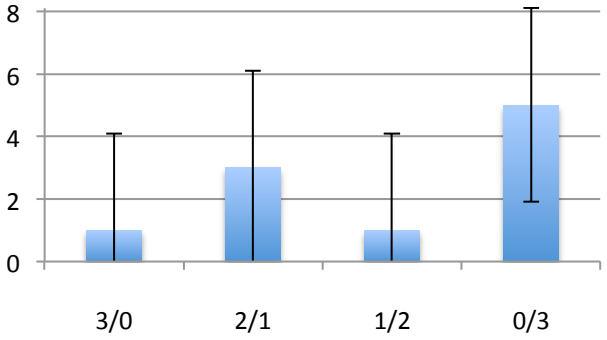
Single-blind or double-blind
(1 author responded)



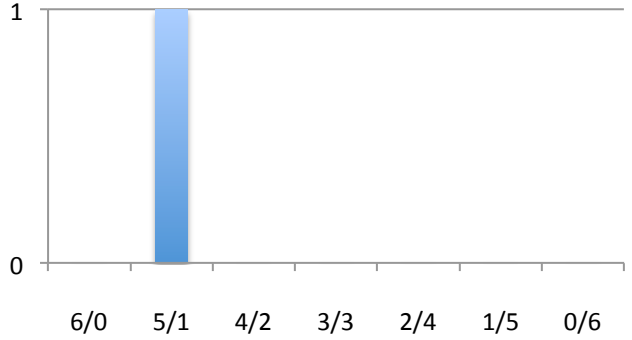
single-blind / # double-blind
(2 authors responded)



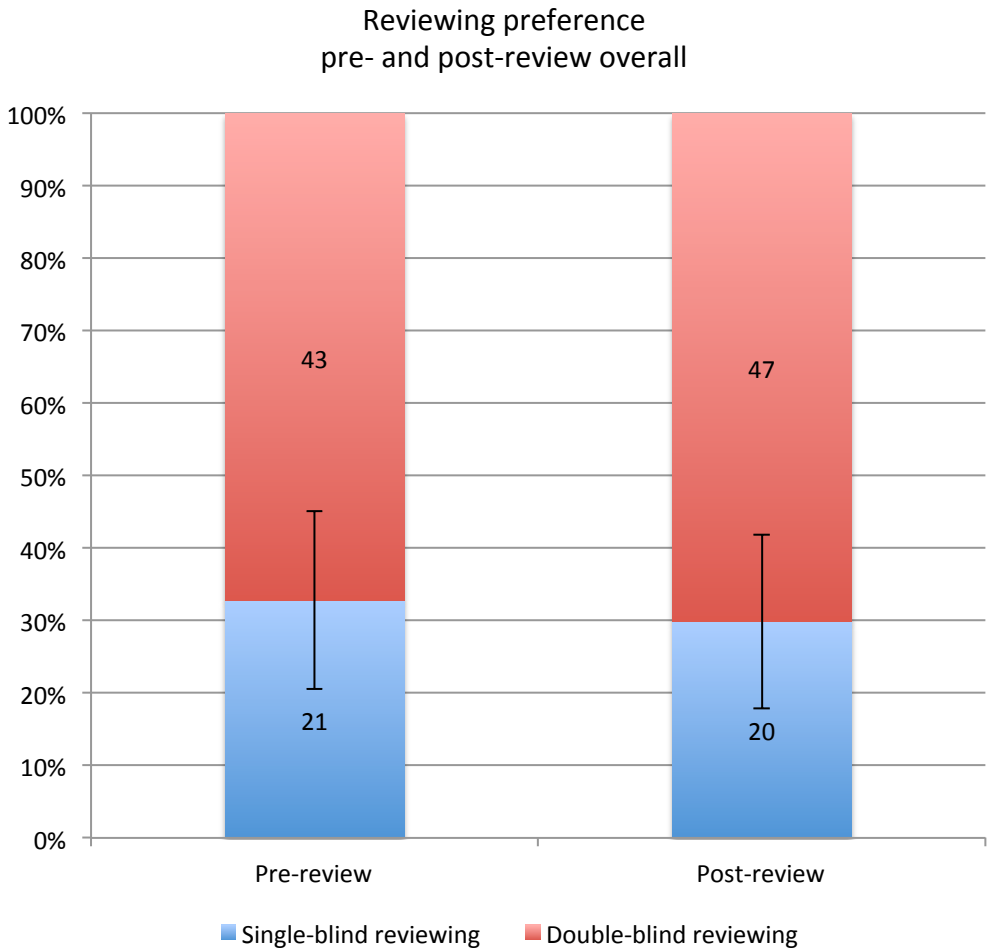
single-blind / # double-blind
(3 authors responded)



single-blind / # double-blind
(6 authors responded)



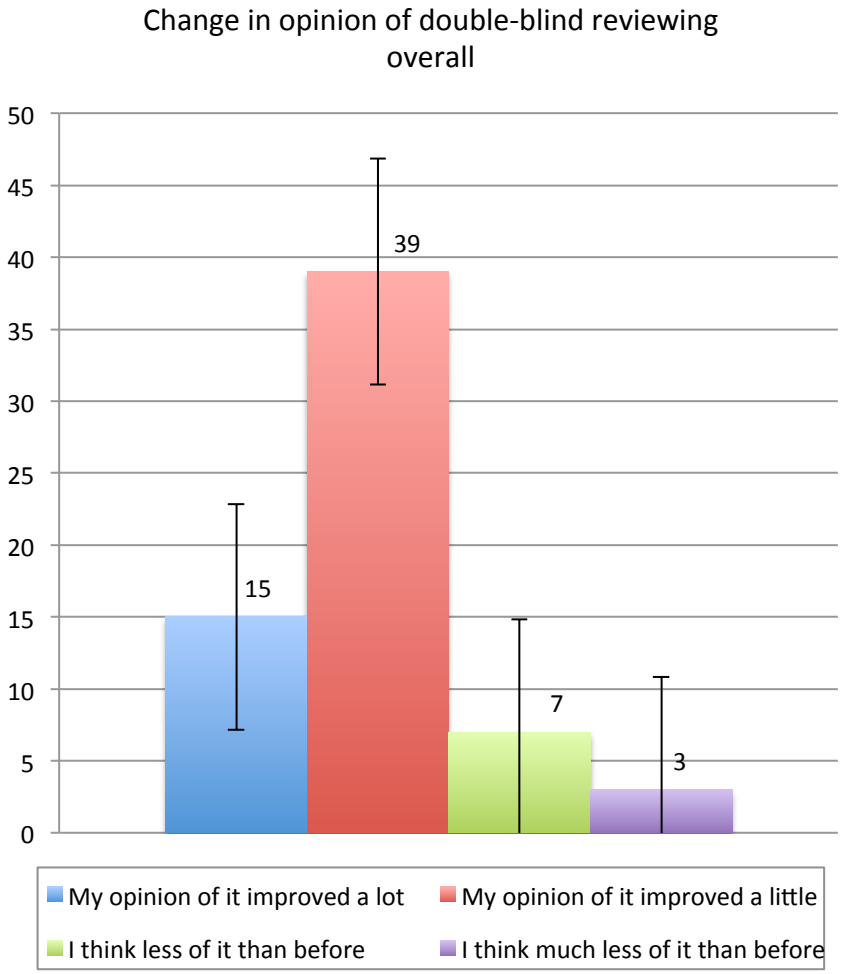
SBR or DBR: committee



Pre-review:
67% +

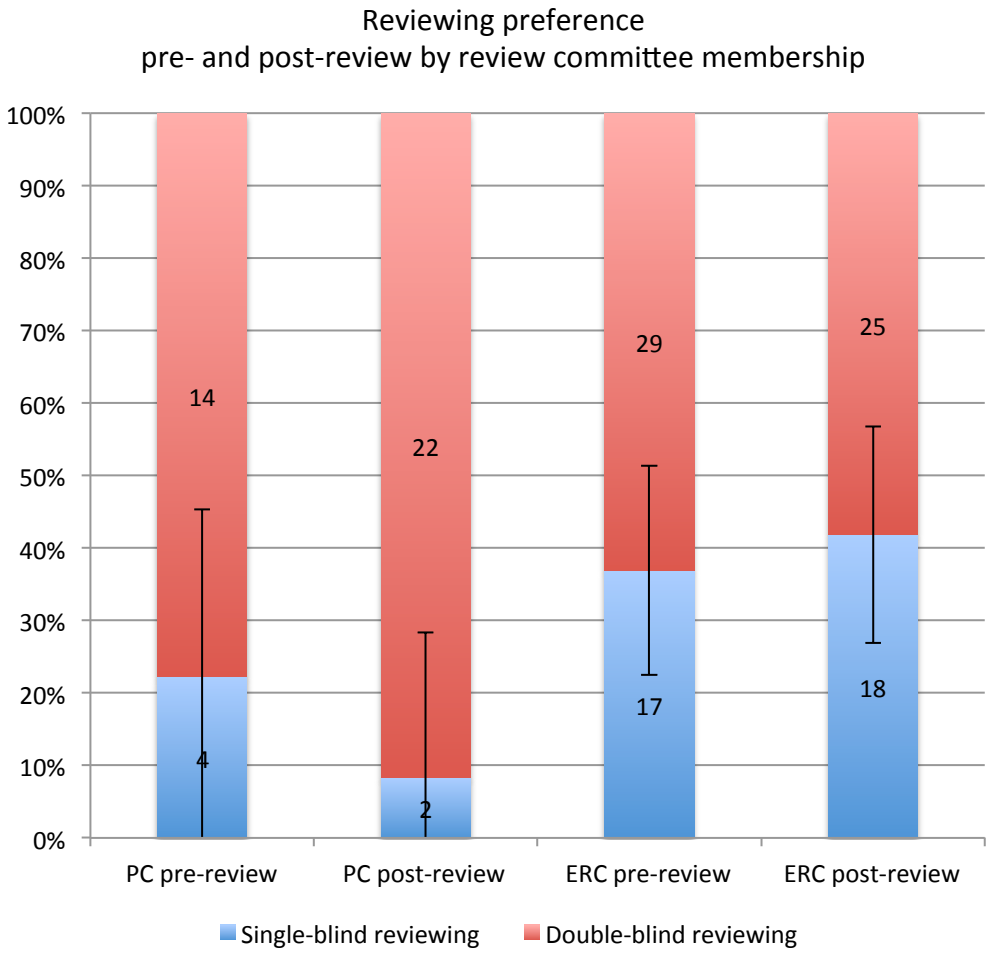
Post-decision:
70% +

SBR or DBR, change, all authors



84% of respondents were more favorable

SBR or DBR, by PC vs. ERC



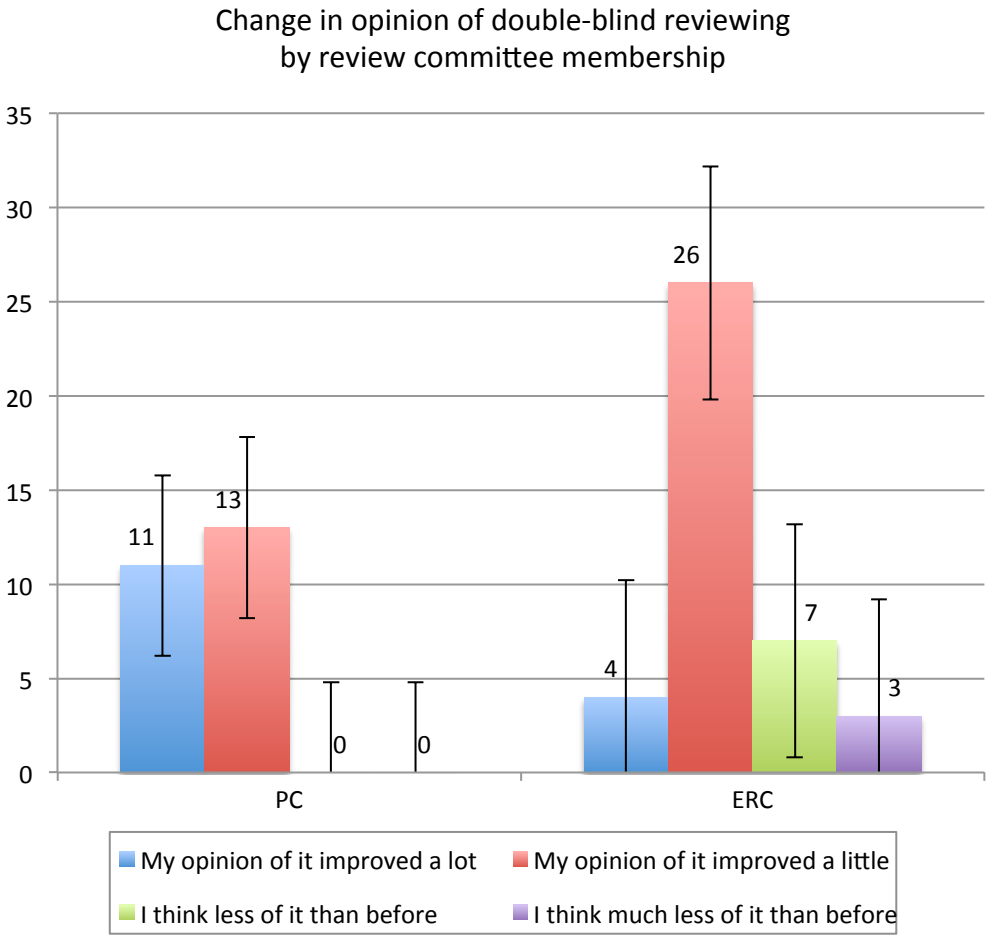
PC

Pre 78% Post 92%

ERC

Pre 63% Post 58%

SBR or DBR change by PC/ERC



PC

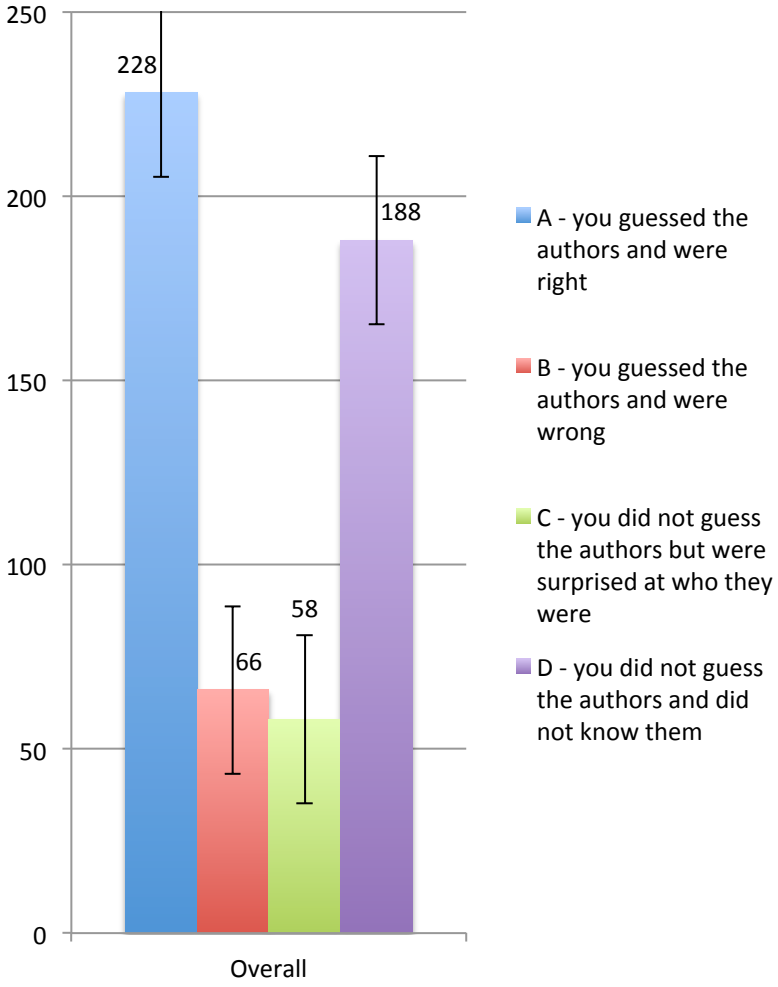
100% more favorable

ERC

75% more favorable (but more against using it!)

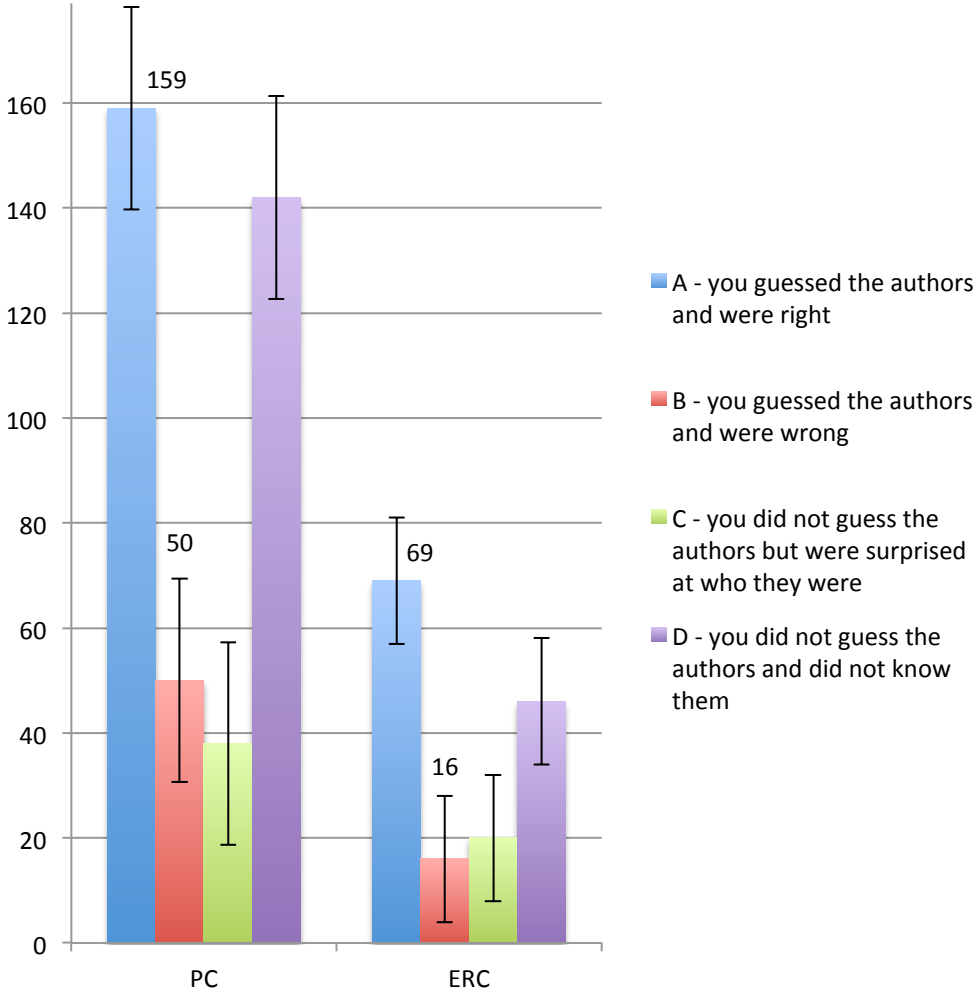
Why this difference between PC and the ERC?

Blinding efficacy, overall



77% guesses were correct

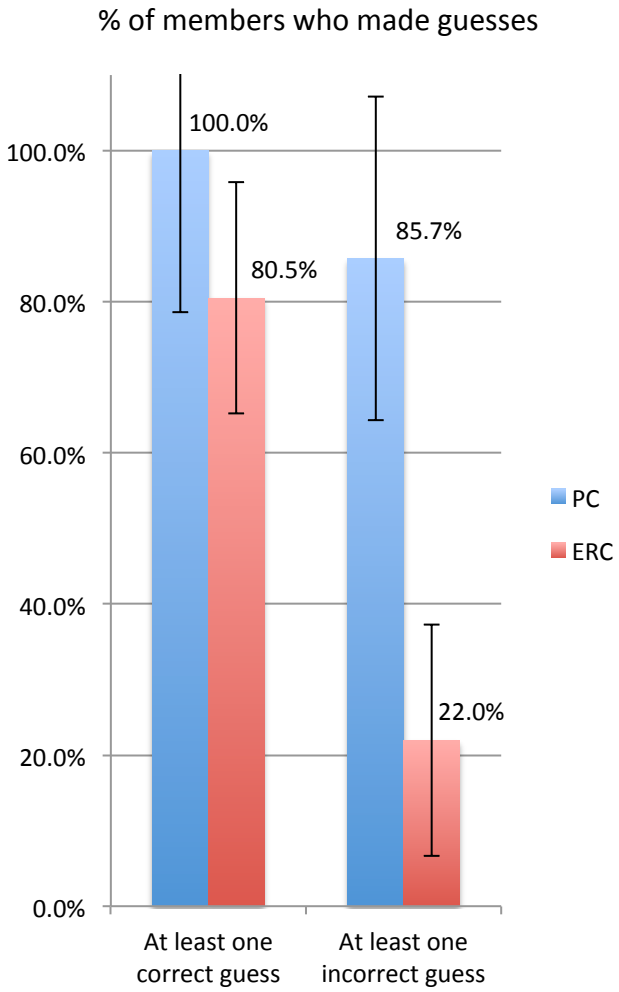
Blinding efficacy, by PC/ERC



PC: 76% correct

ERC: 81% correct

Guessing experiences per PC member



PC: 81% had at least one incorrect guess

ERC: 22% had at least one incorrect guess

Possible explanation: only 1 in 5 ERC members ever guessed wrong, so most failed to see the point of blinding

Anecdotes (PC/ERC)

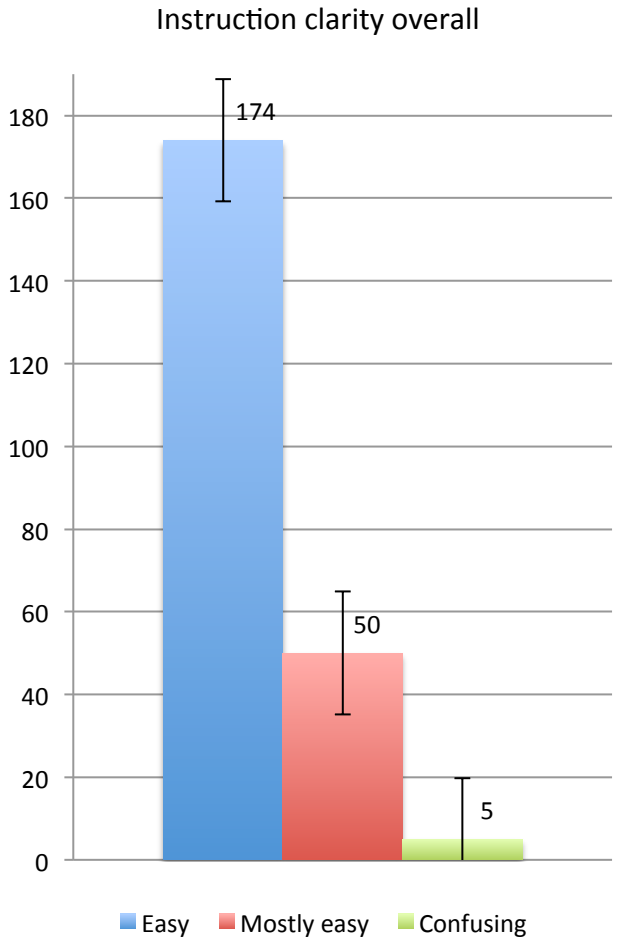
- “On two submissions, if I had known the authors, I would have started with a higher opinion of the submissions than was justified. I would have realized before long, but the DBR saved me the time of realizing that people I respect could have done better.”
- “I was really surprised by authors in a handful of cases. My reviews might have been biased if I'd known authorship up front.”
- “I thought I would be able to easily guess who the authors of various papers were. I turned out to be mistaken in many cases, and as a result, I stopped thinking about authors altogether. I think this is a good outcome.”
- “Despite my paper being rejected, I think DBR was overall a good thing - in one case I found myself tempted to change my opinion of the paper when I knew the identities. I think maintaining blinding throughout the review process would be beneficial (but it is probably not feasible)”

Anecdotes (PC/ERC)

- “I already had a negative opinion of DBR, but my experience as an author, trying to conform to the guidelines while describing work that is part of a larger industrial and academic context, was extremely frustrating. It was not possible to do a really good job, and I believe this actively impeded the reviewers. Meanwhile, as an ERC reviewer it was clear which groups had produced all the papers I saw.”

- “As a reviewer, my opinion improved because there was one instance where I was truly surprised. The degree of surprise suggested to me that I might have held some unconscious bias for the authors had I known their identity in advance. As a writer, I didn't like it because the paper I wrote built upon a previous paper I wrote.”

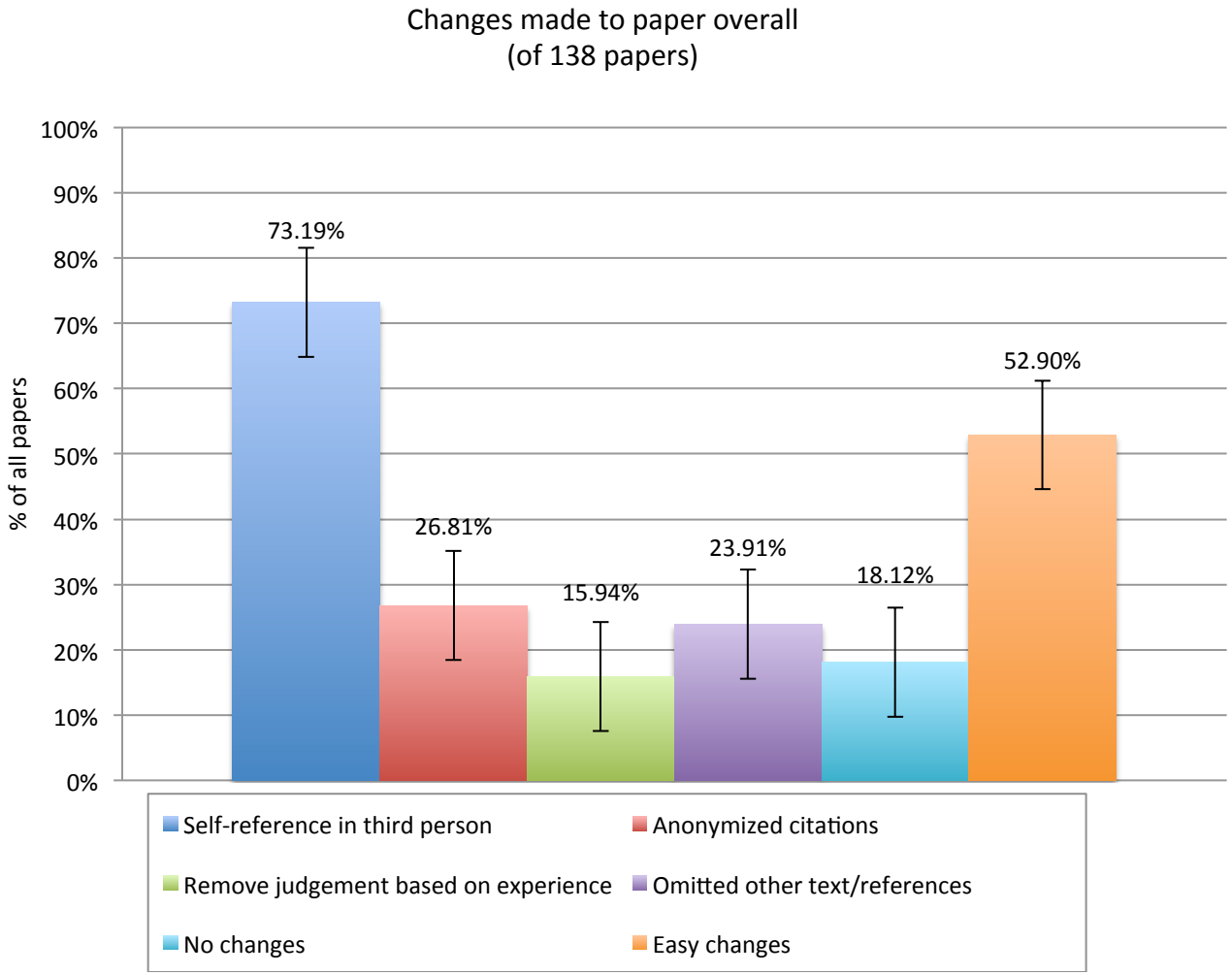
DBR instructions



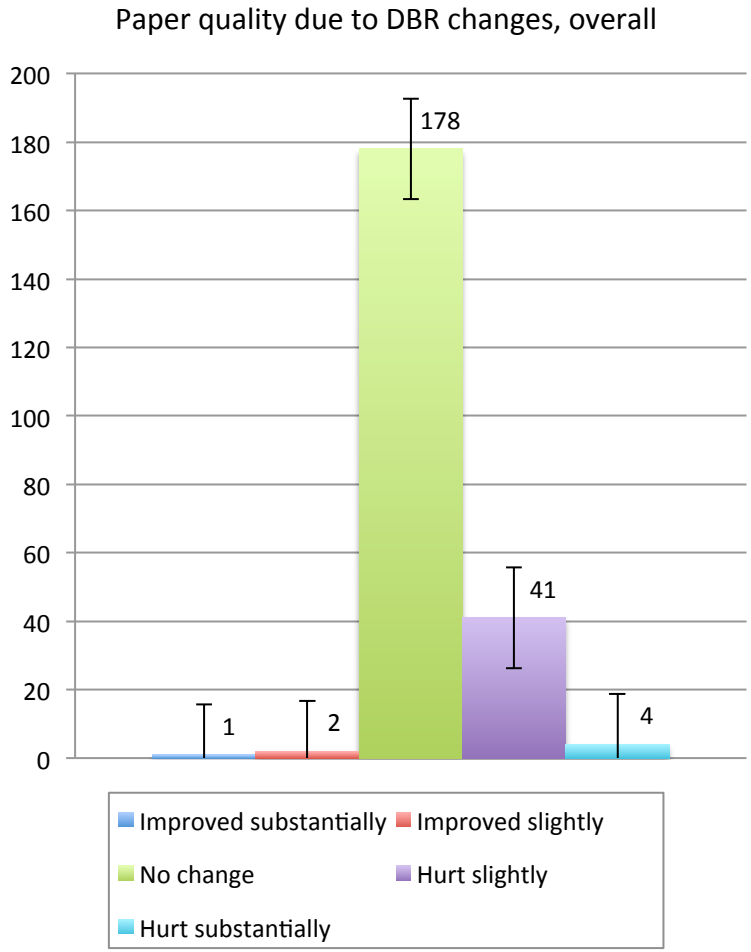
76% found instructions easy

98% found instructions mostly easy, or better

Changes made to paper



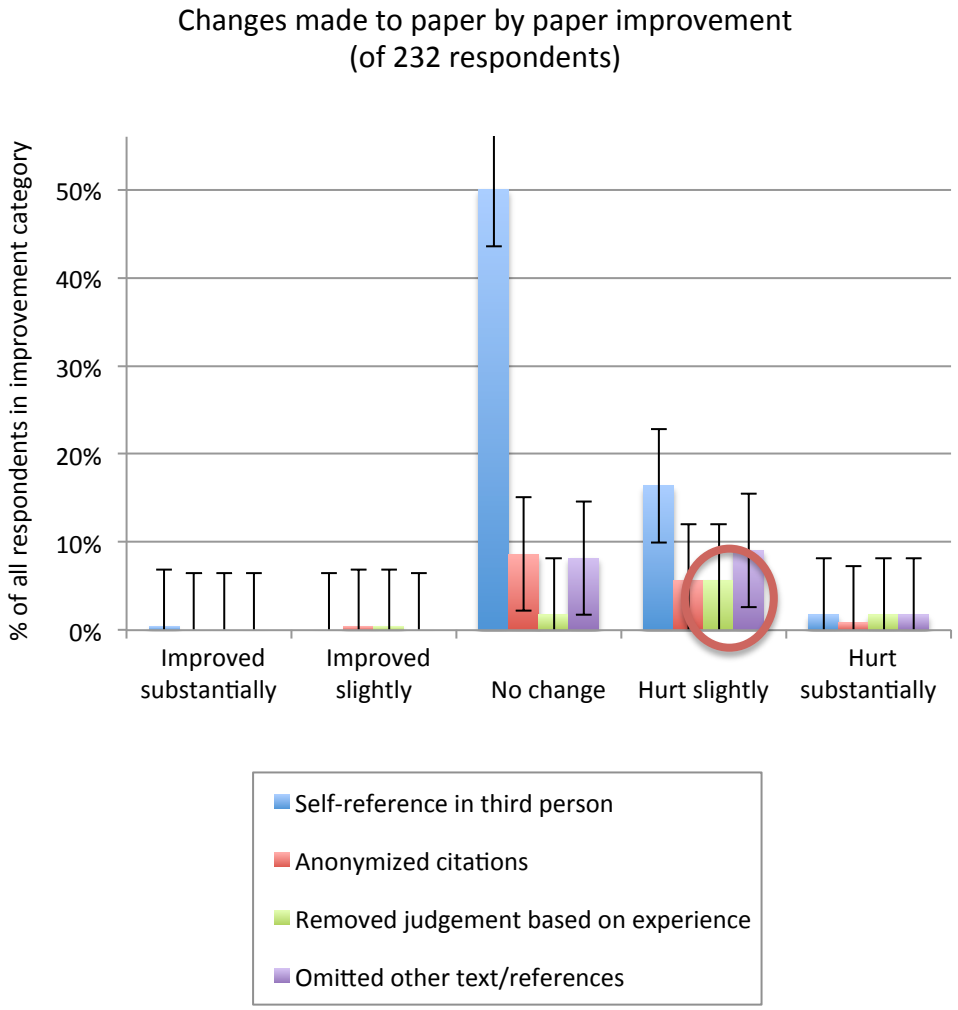
Judgment of DBR changes



80% felt paper not impacted, or improved

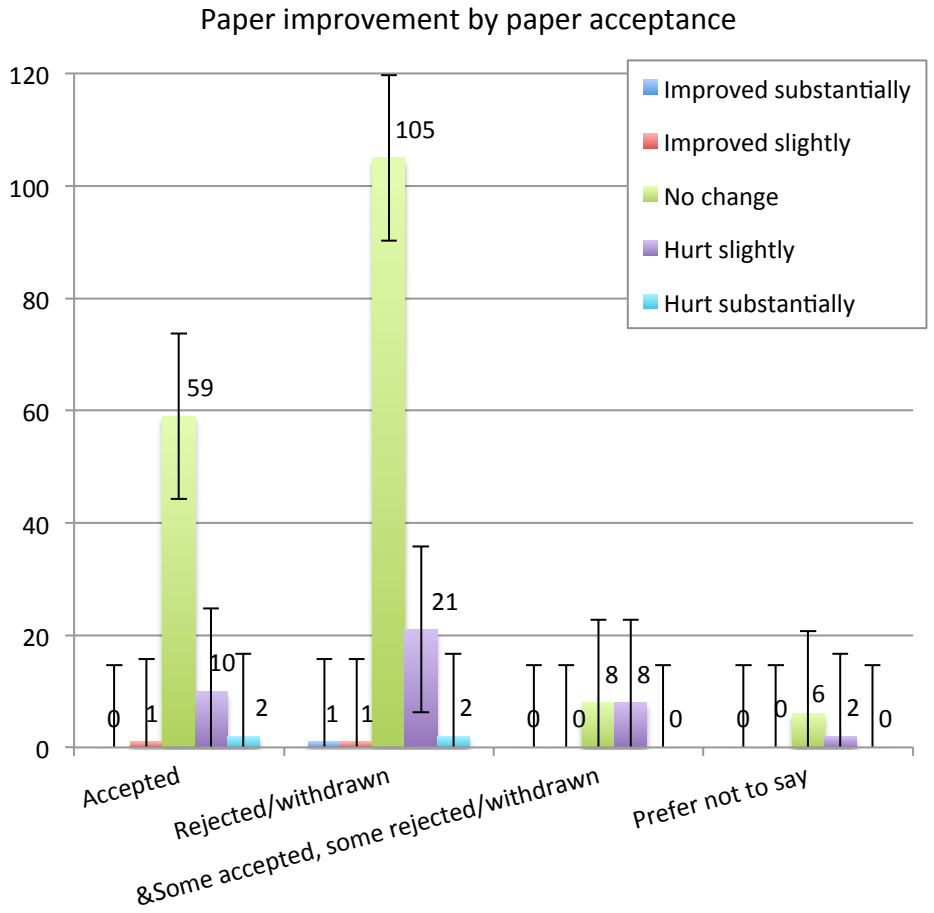
2% felt paper hurt substantially

Correlating change to judgment



Removing experiential assessment viewed as somewhat harmful

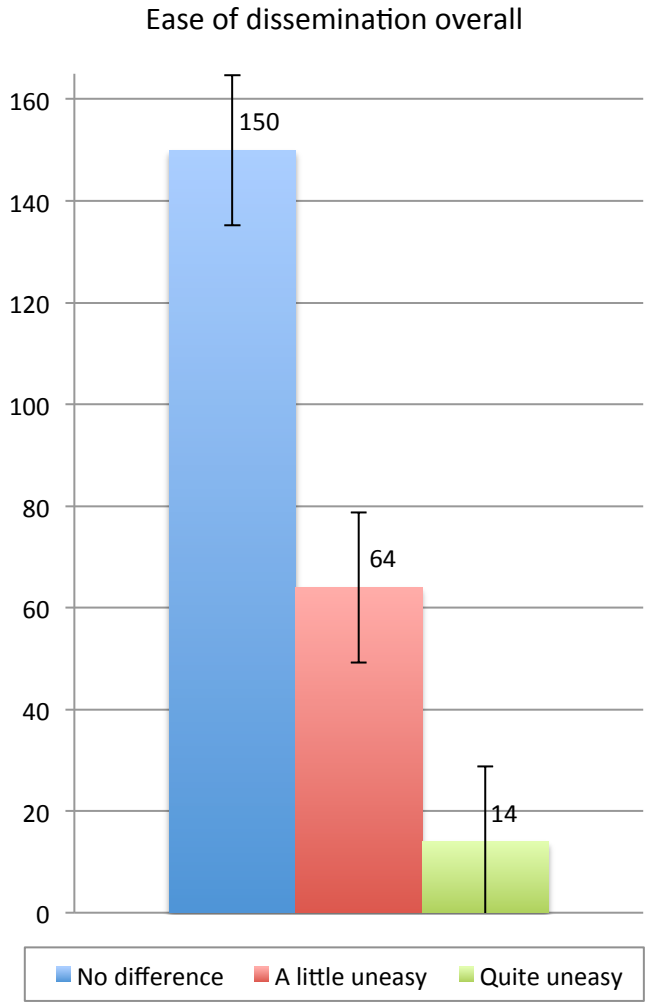
Authors' judgment not correlated with acceptance



18% of rejected papers "hurt"

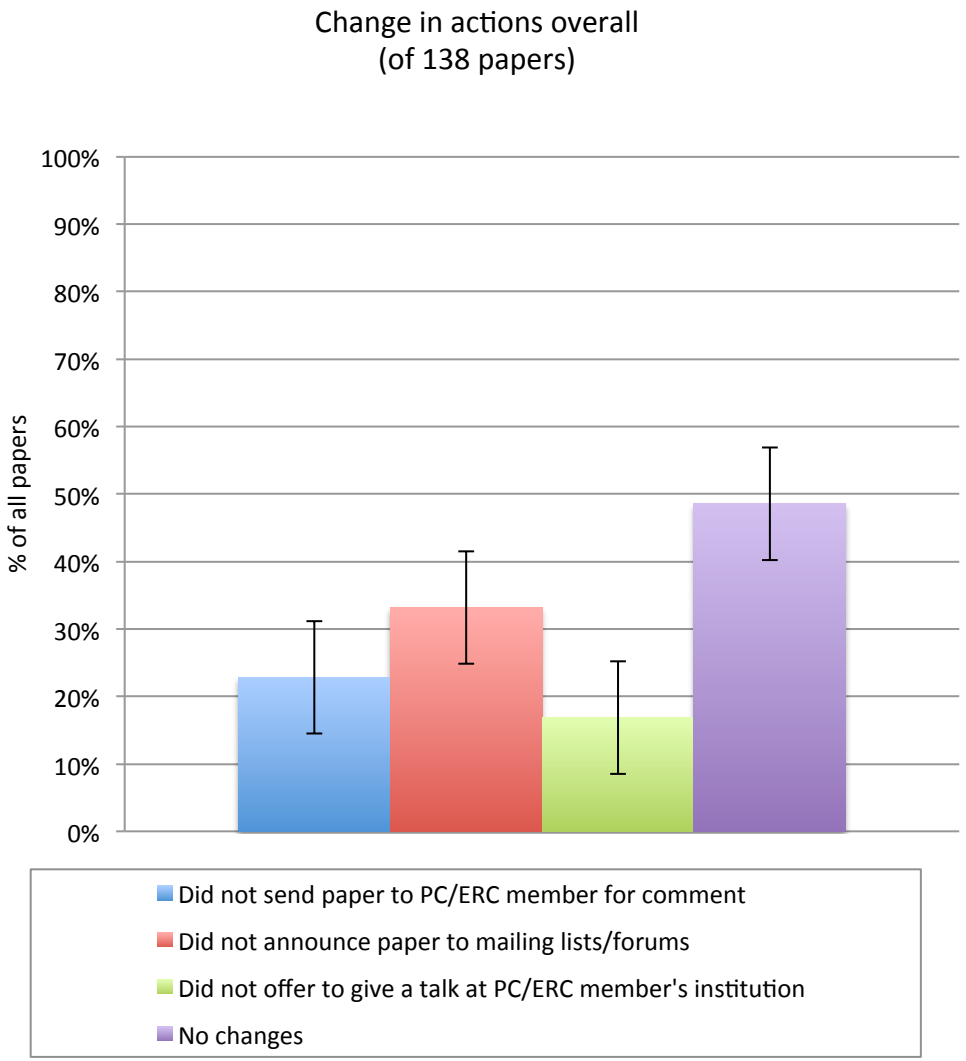
17% of accepted papers "hurt"

Feeling about dissemination



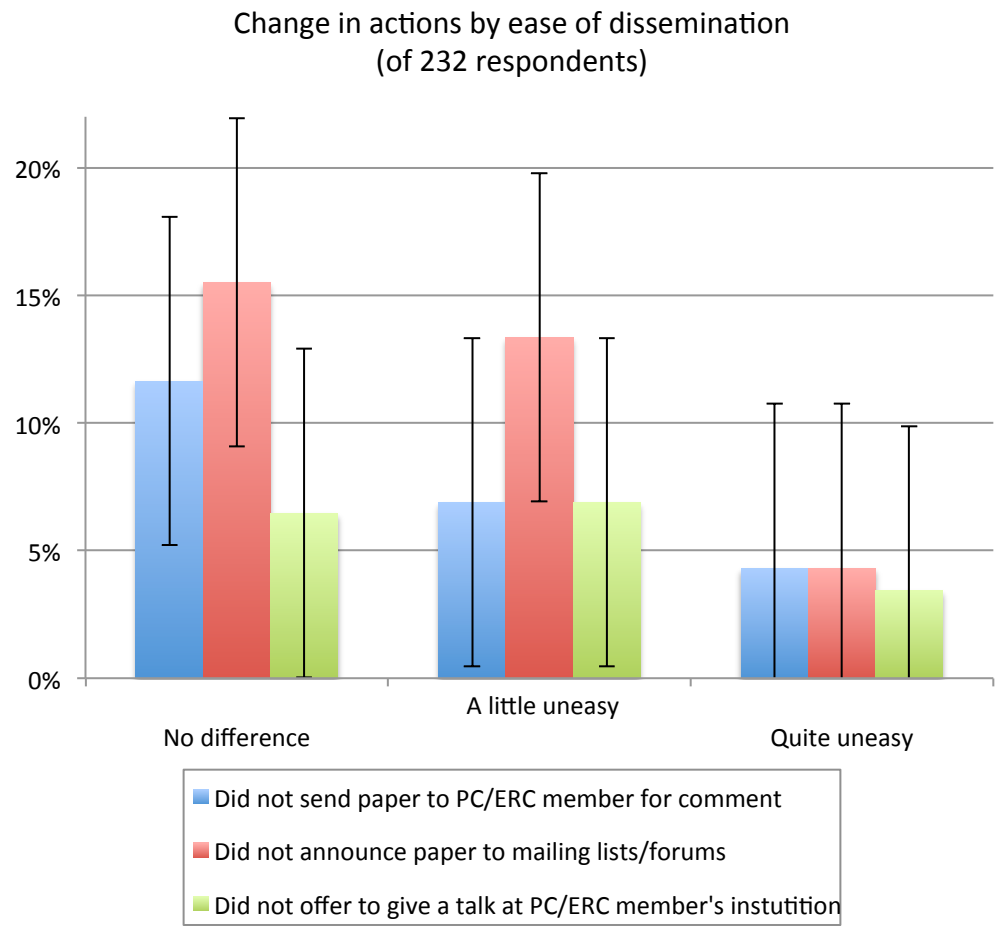
65% indifferent
29% a little uneasy
6% quite uneasy

Change in dissemination behavior

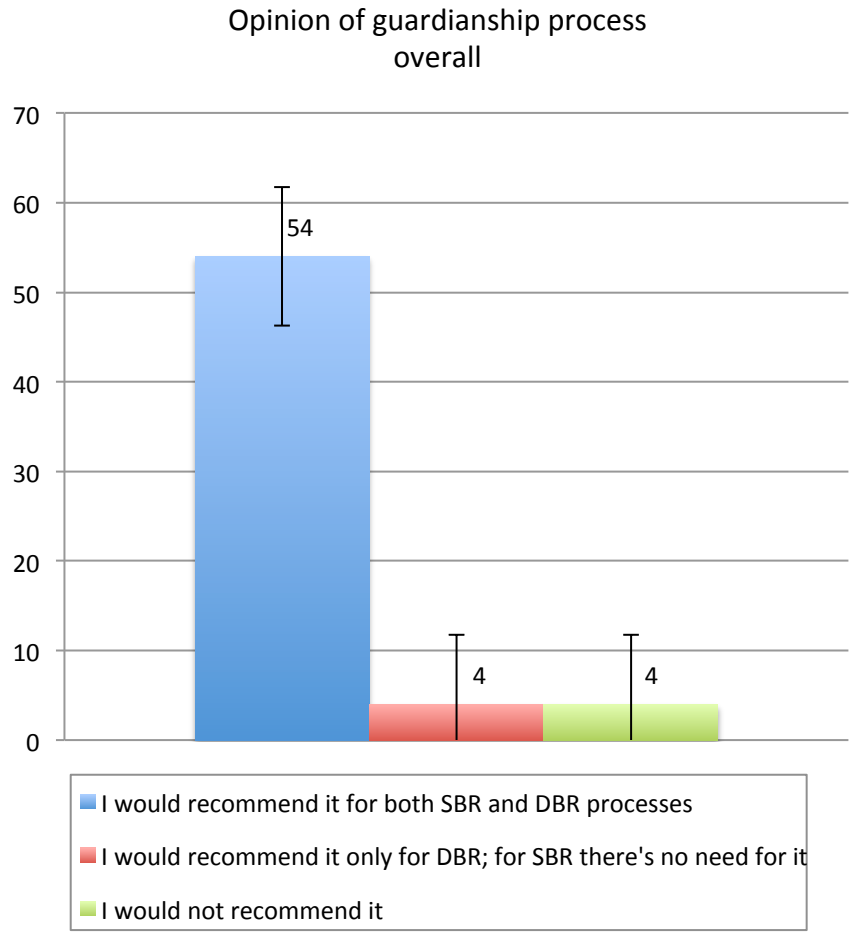


51% of papers' post-submission dissemination actions impacted

Change in behavior broken down by (dis)ease about dissemination



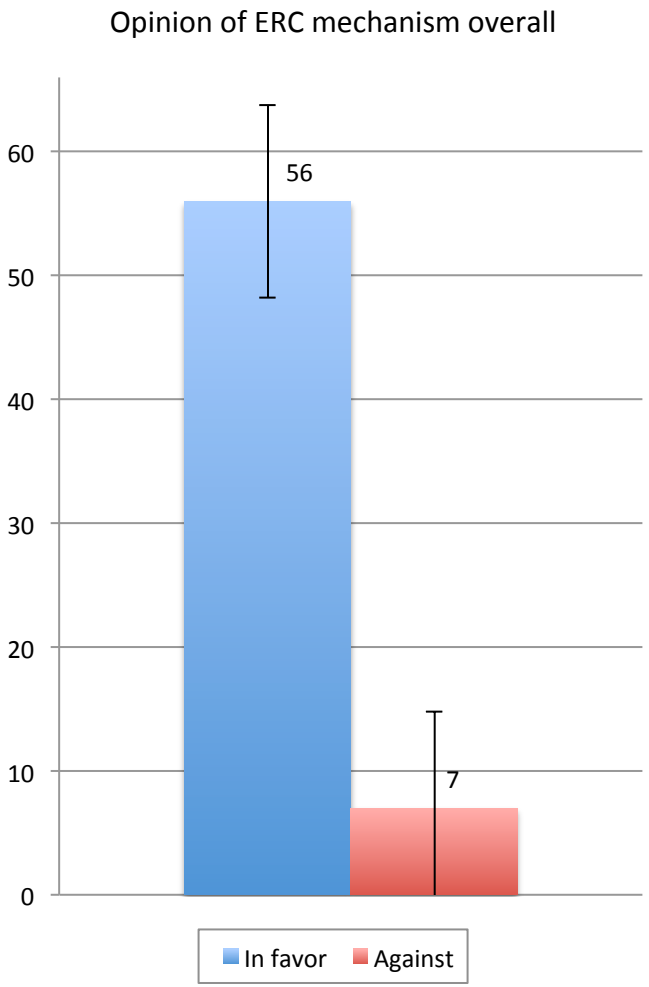
Guardians: good idea?



87% in favor overall

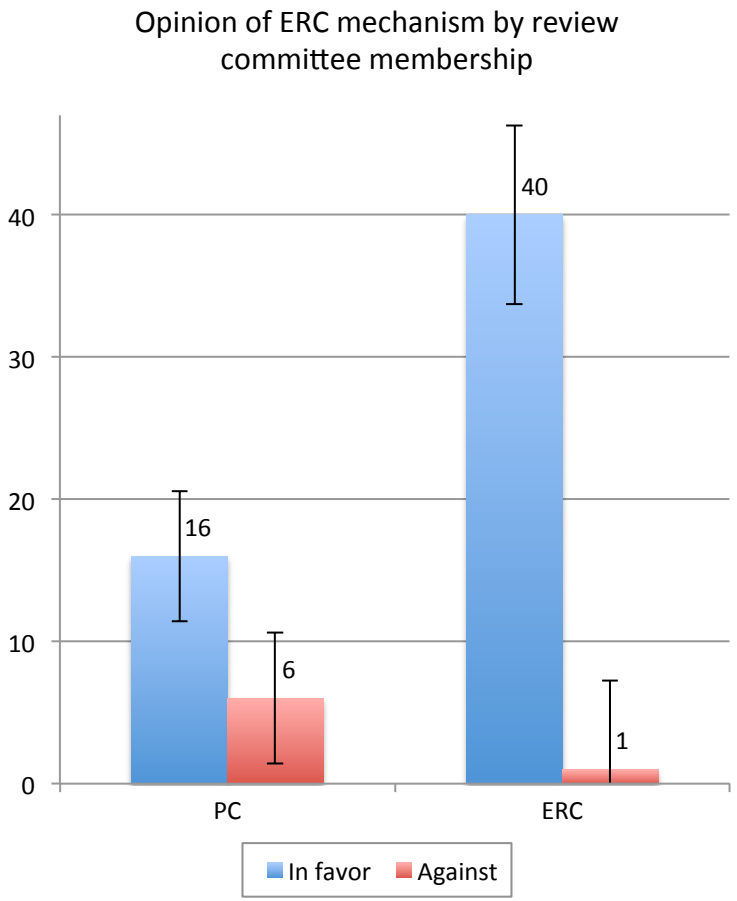
93% in favor for DBR

ERC: to use or not to use?



89% in favor

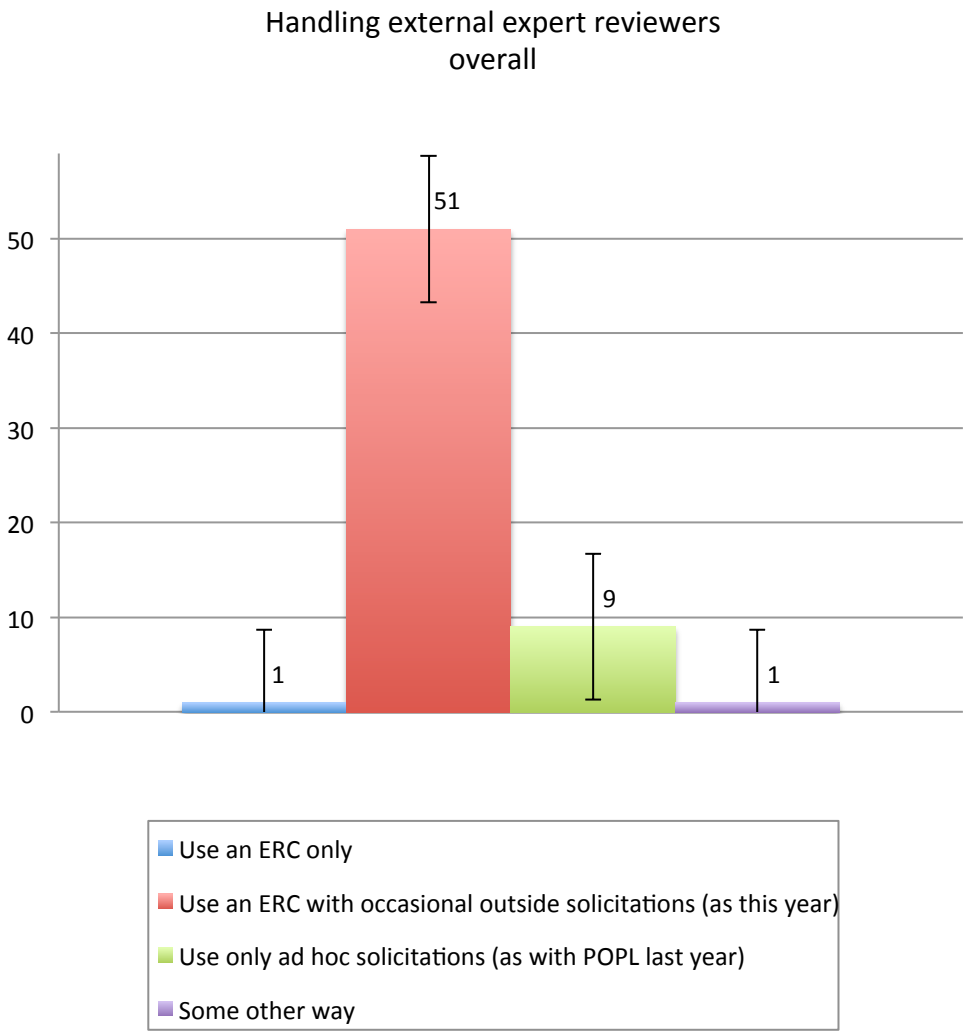
ERC, by committee



PC: 73% in favor

ERC: 98% in favor

Handling expert reviews



ERC+outsiders: 82%

Outsiders only: 14%

Summary: Process judgment

- Light DBR is viewed favorably overall
  - PC very favorable (92%), authors favorable (72%), ERC only slightly (58%)
    - ERC members may be less disposed because they were rarely surprised by authorship (22% vs. 81% of reviewers)
    - Overall blinding success: 23% of guesses wrong (consistent with past studies)
  - Anecdotes: DBR makes a difference
- ERC and Guardians viewed very favorably (89% and 93%, respectively)

Summary: Light DBR costs

- Paper preparation
  - Instructions mostly easy (98%)
  - 47% of papers: non-superficial changes
    - 20% of authors: changes hurtful (4%: a lot)
    - Assessment did not correlate with acceptance
- 51% of papers' post-submission dissemination impacted
  - 35% of authors uneasy about breaking rules

Future Work

- Assess outcomes!
- Problem: hard to do

Many thanks

- The PC and ERC: you amaze me
- John Field and the steering committee
- Other past PC chairs
  - Alex Aiken (PLDI), Andrew Myers (S&P), Todd Mowry (ASPLOS), Kathryn McKinley(PLDI), David Wagner (Sec)
- Eddie Kohler for HotCRP
- Greta Yorsh, for help arranging the program
- The submitting authors: what an amazing set of papers!