Notes from the WiGS 2015 Fall Discussion

Presenter: Carole Lee
Topic: Bias in Peer Review
Date: 10/20/15

Truth and innovation:

- Journals like to publish statistically significant effects
- Pressures authors into fraudulent research practices, increases false positives
- Grant review is thought to be biased against innovative work - certain kinds of proposals are never even submitted
- Broken process - keeps us from our goals

Can we blame commensuration bias?

- Commensuration = taking heterogenous qualities and distilling to a single metric
- It is powerful and useful to be able to compare different types of things
- Commensuration takes place in both peer review and grant review — lots of qualities like innovation, significance, approach, qualifications, institution, broader impacts, etc… and it all gets distilled into some final assessment
- This all becomes a final score along a single scale of value - overall impact score on a 9-point scale, and reviewers define the weighting of the components
- This weighting is a real issue — Rockey 2011, Langfeldt 2001, Lamont 2009 — weightings vary across reviewers

Questions we can ask:

- How should we be doing it?
- How do scientists weigh criteria, and are there better ways?

Commensuration bias in journals:

Journals evaluate:

- novelty - discovery of new truths
- methodological soundness - how likely is something to be true?
- significance - how important is this truth?

What is truly important?

- replication of results
- null results
- but no one publishes null results — everyone wants to publish statistically significant results - authors fail to prioritize these submissions (Ioannidis 1998, Dickersin 192, Easterbrook 1991)
- Really important to represent results honestly especially in clinical publications

- Replications aren't popular either - editors straight up say that they are less important than new results, especially failed replications

Bias for significance and method over novelty

Commensuration bias in grants:

- If you go based on the program statements, should weigh innovation more than method
- BUT opinion and some research indicates that method is more important than innovation (Rockey 2011)
- Scores for approach are way more important
- Highly innovative projects involve methodological risk - unproven, new, or difficult
- Methodological rigor and innovation are polarized

****Lots of individuals' decisions go into creating commensuration bias — need to make changes at levels of reviewers, editors, institution, etc. ****

How can we reduce commensuration bias?

- Why are reviewers not just given an algorithm for the weighting? Canadian system has this
- Could also decrease the number of criteria given to reviewers — some publications are going this route, taking significance out and just asking for evaluation of methodological soundness (encourage null results)
- At the editorial level, can give a second look to submissions that score highly on the less popular metrics - innovation for grants, or methodology for publications
- Could also allot fewer dollars to more proposals
    - Norwegian study — acceptance rate goes up, novelty ratings go up
- At the institutional level — parallel systems with the same branding to give a home to publications that don't score as highly on some metrics (e.g. Nature communications)
- Can also create a less fine-grain scoring system — creates more ties, and then the institution can decide what policy to use as a tie-breaker
- Can also work at the community level - online sharing technologies, public archiving

****Any of these strategies would change what is valued in science — this would be a cultural change****

How might an applicant's race or gender affect this bias?

- If people are able to weigh criteria however they want, then evaluators will prioritize whatever criterion favors the white male candidate (Hodson 2002, Norton 2004, Uhlmann and Cohen 2004)

- Commensuration bias can be measured when white male researchers receiver higher overall scores than minority/female researchers even though they have received identical scores on individual peer review criteria (this is her study idea)
- Do have data showing that black and Asian applicants less likely to get funded by NIH, even controlling for EVERYTHING. 80,000 R01 applications. See the bias in the first round of review.
- Focus on subconscious bias on how that weighting happens going into the overall score
- Book: How Professors Think: Michele Lamont

Discussion:

- Where could she get the data for her study from?
    - GRFP - could our grant-writing class help?
    - By crowdsourcing, would it select for a particular group of people? People who got bad ones?
    - PLoS?