SCIENTIFIC PEER REVIEW: INSTITUTION WORKINGS & OPPORTUNITIES FOR RCT'S

Kevin Boudreau,
Northeastern & NBER

NBER Economics of Science meeting
Sept 25 2025

Overview

- Objective:
  - Perspectives on paths forward for RCTs in scientific peer review
- Scope:
    - To advance research + practice
    - Journal review + grants

Process

1 Introduction 3

2 Historical Context of Scientific Peer Review 7

2.1 Contemporary Forces Continuing to Shape Peer Review 8

3 Institutional Performance—and Questions Raised 10

4 Today's Organization of Peer Review 15

4.1 Peer Review Governance & Structure 16

4.2 Objectives & Submission Selection Criteria 19

4.3 Peer Review Process 23

4.4 Division of Labor & Capital 32

4.5 Incentives, Motivations, and Norms 33

5 Perspectives on Opportunities for RCT's 36

5.1 RCT Opportunities for Institutional "Tinkering" 37

5.2 RCT Opportunities for Advancing Understanding of Theoretical Mechanisms 39

5.3 Potential for Larger Institutional Leaps: Novel Organization & Digital Infrastructure 43

5.4 A Recipe for Realizing RCTs 43

References 47

Appendix: Journal and Grant Objectives 58

Behold Peer Review! A Remarkable Institution

Para-Organization

- Organized outside main employers
- Anonymous (blinded) actors and inputs
- Both highly centralized platform and distributed contributors at once
- Relies on millions of hours of unpaid academic labor

Functions for Science Community

- Select promising papers and proposals
- Certify and validate new knowledge
- Maintain standards and fairness
- Allocate resources and prestige across global community

Internal Institutional Logic

- Protect legitimacy and credibility, incentives
- Operate under tight budgets and limited capacity
- Bathed in norms and culture of Science

Not a "Gold Standard": Recent, Evolving Institution

- Blinded, external review = a 20th-century invention
  - Most of history: evaluation = reputation, lineage, performance (Arab, European, Asian traditions)
  - Early 20C: editorial discretion + "gentlemenly" consultation
    - 1936 Albert Einstein & Nathan Rosen episode: gravitational waves to Physical Review ... (Kennefick 2005).
  - Mid-C: big science, output, funding, specialization, legitimacy
  - 1973: Nature journal finally shifts to external, anonymous review

- Recent pressures: Digitization, online platforms, int'l submission growth, political/policy pressure, open, automation/AI, etc.

Einstein, A. & Rosen, N. (1937) 'On gravitational waves', *Journal of the Franklin Institute*, 223(1), pp. 43–54.

Performance: Literature Takes Dim View

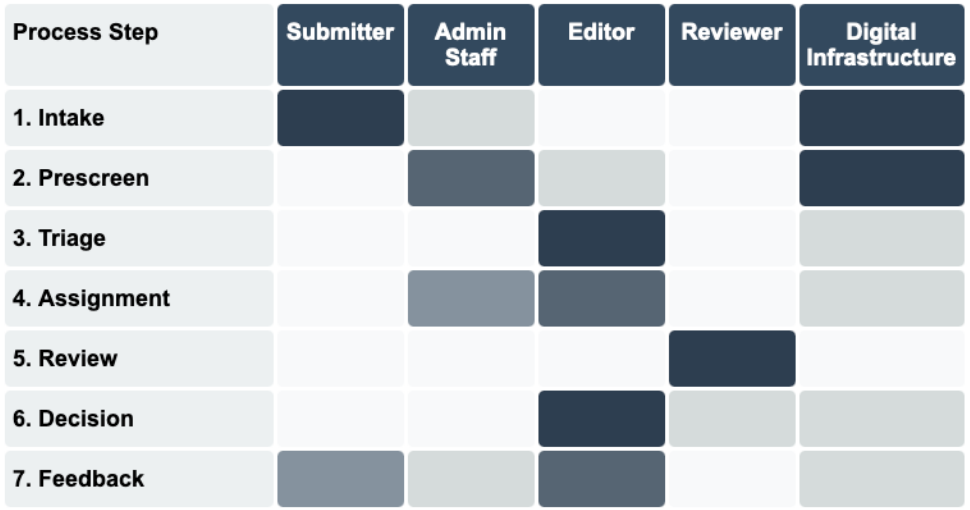
(My perspective: We need better understanding)

- "Reliability" (differences across evaluators) ✅ ❓
- "Bias" (correlations between submissions/submitters & evaluation) ❓
  - Peters & Ceci 1982; Blank 1991; Tomkins et al. 2017; Witteman et al. 2019; Boudreau et al. 2016; etc.
- Burden: 60M+ hours annually, concentrated in a minority of scholars ✅
- Recent public "credibility crisis" High-profile retractions, replication failures, public scrutiny
- But:
  - Limits to interpretation: descriptive patterns, limited theory, discerning tests
  - Peer review = endogenous institutional response (better alternative?)
  - Evidence: discerning bad from good, predicting longer run outcomes
  - Limited theory, theoretical benchmarks, empirical benchmarks

First Half of Chapter: Institutional Architecture: Journals + Grant Processes

- 4 Today's Organization of Peer Review
  - 4.1 Peer Review Governance & Structure
    - 4.1.1 Journals
    - 4.1.2 Fund Granting Bodies
  - 4.2 Objectives & Submission Selection Criteria
  - 4.3 Peer Review Process
    - 4.3.1 Submission Intake & Requirements
    - 4.3.2 Pre-screening
    - 4.3.3 Initial Rejections and Triage.
    - 4.3.4 Reviewer Selection and Invitations.
    - 4.3.5 Review and evaluation.
    - 4.3.6 Deliberation and judgment.
    - 4.3.7 Feedback and iteration.
    - 4.3.8 Verification Processes prior to Final Acceptance
  - 4.4 Division of Labor & Capital
  - 4.5 Incentives, Motivations, and Norms

Division of Labor + Capital



Four Perspectives on Opportunities for Peer Review

(Drawing on the preceding detailed characterization)

- 5 Perspectives on Opportunities for RCT's
  - 5.1 RCT Opportunities for Institutional "Tinkering"
  - 5.2 RCT Opportunities for Advancing Understanding of Theoretical Mechanisms
  - 5.3 Potential for Larger Institutional Leaps: Novel Organization & Digital Infrastructure
  - 5.4 A Recipe for Realizing RCTs

1. RCT Opportunity: Institutional "Tinkering"

Many opportunities for RCTs..

- Even w/o much theory...
- "Economist as plumber" (Duflo, 2017)
- Vaste white space for progress in research + practice
- Scarce resource: getting organized (see Perspective 4)

Institutional Dimension	Illustrative Examples of Plausible Treatments
Governance & Structure (Section 4.1)	Presenting journal identity differently (e.g., society vs. commercial framing in instructions); signaling whether editors are professional staff vs. academic volunteers; varying transparency about editorial board composition.
Objectives & Criteria (Section 4.2)	Adjusting weighting of novelty vs. feasibility in review rubrics; including or removing broader impacts criteria; emphasizing interdisciplinarity in reviewer prompts.
Submission Intake & Requirements (Section 4.3.1)	Varying data/code availability requirements; requiring registered reports vs. standard submission; altering submission fees or formatting demands.
Pre-screening (Section 4.3.2)	Randomizing AI-assisted plagiarism checks vs. human-only checks; varying threshold for similarity flags; signaling stricter vs. lighter COI verification.
Initial Rejections & Triage (Section 4.3.3)	Varying desk-reject thresholds (tight vs. loose); testing AI-assisted triage recommendations; altering transparency of triage rationale provided to authors.
Reviewer Selection & Invitations (Section 4.3.4)	Randomizing assignment by intellectual distance; varying number of reviewers; including vs. excluding author-suggested reviewers; algorithmic recommender vs. editor judgment.
Review & Evaluation Mode (Section 4.3.5)	Randomizing between narrative vs. structured scoring rubrics; single- vs. double-blind; requiring signed vs. anonymous reviews; exposing reviewers to others' reports before submission.
Deliberation & Aggregation (Section 4.3.6)	Varying whether panels use collective scoring vs. rank-ordering; random tie-breaking (lotteries); calibration exercises vs. no calibration; AI-assisted summaries of reviewer input.
Feedback & Iteration (Section 4.3.7)	Publishing anonymized referee reports vs. not; structured vs. unstructured editorial letters; limiting vs. extending R&R rounds; providing author response templates.
Verification Processes (Section 4.3.8)	Randomizing use of statistical checks (e.g., statcheck); requiring vs. not requiring code execution; applying image forensics vs. skipping; flagging citations to retracted work.
Division of Labor & Capital (Section 4.4)	Varying reviewer workload allocation (2 vs. 3 reviewers per paper); testing light honoraria or service credits; providing vs. withholding AI-assist tools for editors; offering calibration training vs. none.
Incentives, Motivations & Norms (Section 4.5)	Recognition tokens (Publons credit, ORCID badging); randomized open vs. closed review identity; adding learning-oriented feedback (sharing co-reports); signaling reciprocity expectations ("your turn to review").

2. RCT Opportunity: Advance Theoretical Understanding

How does Peer Review differ from a simple benchmark?

- Even w/o much theory...
- Abstract peer review as:
  - Information processing + Multi-Criteria Decision Making (MCDM)
  - Quality, Novelty, Feasibility, other
- Simple Benchmark: MCDM in Decision Sciences
  (Keeney & Raiffa, 1976)

"Simple" Problem Characterization

- Prediction (scoring characteristics)
  - Well-defined choice set
  - Known evaluation criteria
  - Complete information (esp. scores)
  - Quantifiable uncertainty

▼

- Judgment (selecting options)
  - Known, stable objective function
  - Known, compensatory tradeoffs
  - Rational optimization

2. RCT Opportunity: Advance Theoretical Understanding

Isolating, chipping-away at mechanisms...

Evaluator/Reviewer Scoring

- Incentive Alignment & Adherence: paradigms, preferences, networks, private incentives...
- Weighting: Judgement mixing with prediction
- Attention & Bounded Cognition:
  - Knightian, Arrow uncertainty
  - Suboptimal, vague, complex rubrics
  - Vague, overlapping, non-ind't criteria
  - Differences expertise, bounded cognition

Editorial/Committee Decision-Making

"Simple" Problem Characterization

- Prediction (scoring characteristics)
  - Well-defined choice set
  - Known evaluation criteria
  - Complete information (esp. scores)
  - Quantifiable uncertainty

▼

- Judgment (selecting options)
  - Known, stable objective function
  - Known, compensatory tradeoffs
  - Rational optimization

2. RCT Opportunity: Advance Theoretical Understanding

Many opportunities for RCTs...

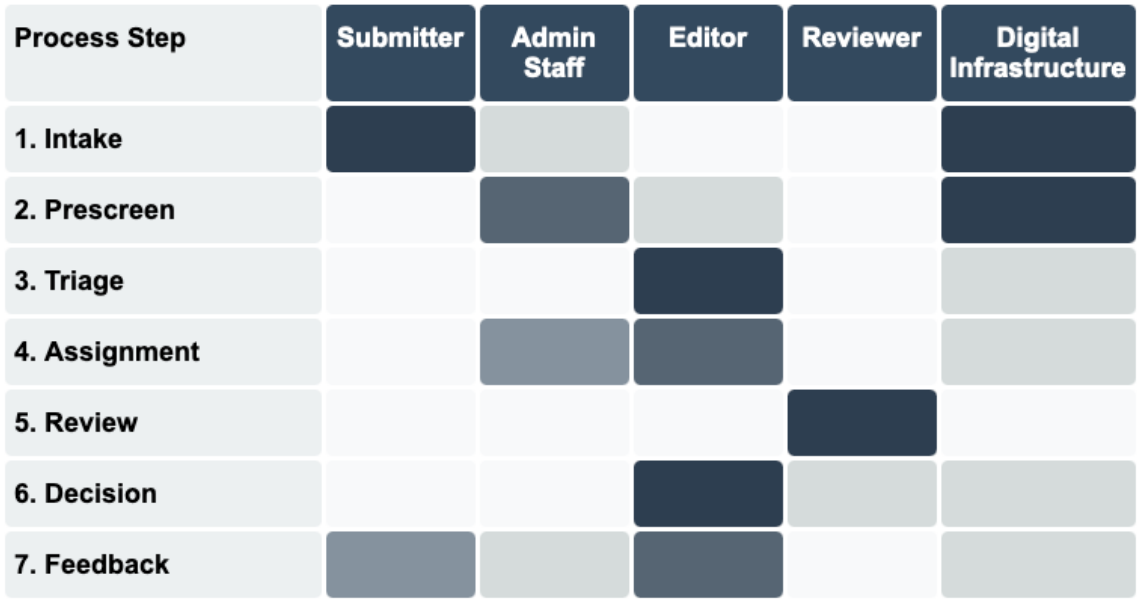
Mechanism / Issue	Prediction & Scoring (Evaluator Role)	Judgment & Decision-Making (Editorial / Panel Role)
Misalignment & Incentives	private vs. public interests (advancing networks vs. advancing knowledge) [blinding identity cues, conflict-of-interest disclosure], monetary vs. non-monetary motivations (prestige vs. financial reward) [reviewer payment, recognition credits, public acknowledgment], effort incentives (high vs. low investment) [effort prompts, gamification, time allocations]	strategic vs. epistemic goals (reputational safety vs. advancing frontier work) [portfolio constraints, randomized emphasis on novelty vs. feasibility], institutional vs. field objectives (disciplinary coherence vs. interdisciplinarity) [explicit vs. implicit goal statements], public accountability vs. private deliberation [open review formats, transparency experiments]
(Non)independence of Prediction/Scoring and Judgment	reviewers imposing implicit trade-offs (feasibility as gating, novelty as compensatory) [bundled vs. separated criteria, weighting instructions], score reinterpretation at the individual level [anchoring examples, calibration feedback]	panels re-weighting or discounting reviewer scores [aggregation rules: mean, median, trimmed mean], outlier handling and second-guessing [explicit outlier protocols vs. editor discretion], portfolio balancing (topic mix, demographic equity) [randomized portfolio targets, constrained optimization dashboards]
Bounded Rationality & Cognitive Mechanisms	individual cognitive load (reviewer overload, time pressure) [workload variation, deadlines, reviewer task size], heuristics in scoring (legibility bias: clarity, CV signals, verbosity) [rubric simplification, exemplars, heuristic prompts], knowledge limitations (expertise vs. unfamiliar domains) [assignment algorithms, expertise matching, reviewer training]	group cognition limits (bounded attention in panels) [structured comparison tasks, explicit trade-off exercises], social heuristics (conformity, seniority effects) [anonymous voting, order-of-speaking randomization], folk theories ("top journals = best work") [alternative framing of objectives, randomized evaluative criteria]

3. Larger Institutional Leaps: Novel Organization, Digitization, AI

Many opportunities for RCTs...

- Humans vs. capital
- Substituting elements or transformation?
- Paths forward via experiments?

Division of Labor + Capital



4. A Recipe for Realizing these Many Opportunities

- Much written on RCTs
  - (Harrison & List, 2004; List & Rasul, 2011; Glennerster & Kremer, 2011; Banerjee & Duflo, 2017; Glennerster & Karlan, 2017; Deaton & Cartwright, 2018; Duflo, 2020)
- Focus here:
  - Translating to peer review
- One approach here…
  - In situ / partnering with review organization
  - Identifying the moment and context
  - Intensive use of IT infrastructure
  - Managing fixed costs + risks

Step	Practical Actions
1. Identify and define a sampling frame	Find high-potential sites where many observations occur in a bounded moment (e.g., grant cycle, submission pipeline). Define the unit of observation (manuscript, grant application, reviewer–assignment pair). Specify the risk set exposed to randomization.
2. Align with the host institution	Ensure the question addresses host priorities (fairness, timeliness, workload). Frame interventions as natural extensions of practice. Secure leadership buy-in by clarifying institutional as well as scholarly benefits.
3. Anticipate and plan for managing relationships and operations	Build trust and ongoing relationships with journals/funders. Identify staff roles critical for implementation (administrators, technical staff, editors). Anticipate the need for relationship management, administration, and operational leadership alongside research design and writing.
4. Secure or build infrastructure	Confirm access to digital/physical systems (submission platforms, databases, panel meetings). Partner with staff who control existing systems, or build add-ons to enable observation, randomization, and treatment delivery.
5. Careful research design	Before operationalizing, clear inference needs to be designed, around the many challenges. This step however is subordinate to earlier steps, which are generally more difficult and constrained relative to research designdtha is usually more room to think your way out of design challenges than to recover from operational missteps. Use minimum interventions (light-touch changes, not workflow overhauls). Define a minimum inference strategy that tests. Research design should abide by princples named earlier: minimum intervention, heigthened feasibility, minimum measures, minimum inferential strategy, abidine the three unavoidable problems of inference
6. Manage risks and costs	Effective divison of labor, minmum intervention, multiple orthogonal interventions, ideally recover investmetns across protracted relationship

4. A Recipe for Realizing RCTs - Some Highlights

- Unintuitive order of operations:
  - Prioritize choosing the sampling frame…
    - (research setting/timing/set-up)
  - Prioritize access/capabilities in digital infrastructure …
  - Align with host institutions on agenda…
  - Minimum intervention, maximal impact…
- 3 immutable problems to face in research design:
  - True research quality never observed
  - Submitter+submission characteristics never randomly assigned
  - Cognitive processes never truly observed

Thank you.