Collusive and Adversarial Replication

Research and Politics
XX(X):1–15

Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Adam Bouyamourn[*]

Abstract

We describe a game in which social ties between members of a research community may discourage prospective replicators from debunking papers that misreport results. In this game, replication is an entrance decision, as a Replicator chooses whether or not to Replicate a given paper. We show that a high level of social connectedness between members of a research community increases the field-wise False Discovery Rate, a measure of the social welfare associated with a healthy publication process. The moral is that larger, more diverse academic fields with fewer social ties are more likely to have an adversarial culture around replication, and that this improves social welfare. We study three proposals to improve replication practices: Random auditing, or police-patrol replication; automated unit tests; and a recent proposal to lower the threshold for statistical significance. We argue that random auditing and automated unit tests can improve social welfare, but that the effect of lowering the statistical significance threshold is ambiguous.

Introduction

A succession of replication crises across multiple disciplines has highlighted weakness in academic safeguards (Ioannidis 2005; Francis 2013; Collaboration 2015). Poor academic practices, including misreporting, p-hacking, model and site selection may or may not be caught by peer reviewers (Leamer 1974; Gelman and Loken 2013; Allcott 2015). Replicability is also tied to the external validity of a phenomenon, and cumulative learning about a process of fundamental interest: if a finding is generalizable, it can be replicated across multiple contexts and settings (Jin et al. 2023; Slough and Tyson 2024). Further, the incentive schemes that determine how poor academic practices can persist, and the optimal design of incentive schemes to prevent these practices, is an area of growing interest (Berinsky et al. 2021; Bates et al. 2022, 2023).

We study replication in an *entrance* model, in which replicators can choose to opt in to performing replications. We highlight *the role of social connectivity in the publication process*, which is a source of bias that is infrequently discussed in the literature. In our model, preference alignment between researchers and prospective replicators is a key determinant of whether or not replication actually occurs, and the reliability of the research that results.

Preference alignment between agents matters in auditing contexts (Anastasopoulos and Asteriou 2019), and when considering the organization of publication incentives (Fourcade 2009). Further, cultural and social factors can be modelled as supervenient features of strategic interactions between individuals (Lewis 1969; Bednar and Page 2007; Stirling and Felin 2013; Patty and Penn 2020).

We show that a high degree of preference alignment between researchers and replicators can induce replicators not to replicate a given paper, which leads to increases in the *field-wise False Discovery Rate* (FDR) (Benjamini and Hochberg 1995; Storey 2003), as examples of misreporting go unchecked. We show that the FDR is a measure of social welfare, as increases in the FDR harm the reading public. One moral is that lowering the cost of replications is not sufficient to ensure that replication occurs in equilibrium: incentive-compatibility also matters.

In an extension, we model the impact of random auditing, or *police-patrol replication*, and show that it reduces the FDR. We also generalize our results to the case where replications are imperfectly informative, and find that there is a threshold level of informativeness above which our argument applies: if replications are not sufficiently informative, no learning occurs from research in equilibrium.

We study the possible impact of a recent proposal to reduce the significance threshold from $p \leq .05$ to $p \leq .005$ (Benjamin et al. 2017). We show that lowering the p-value threshold has an ambiguous effects on researcher and replication incentives: in our model, it can incentivize misreporting, by making it harder to publish research, but it *also* encourages prospective replicators to enter the market for conducting replications as the opportunity to debunk papers increases. We also discuss automated acceptance testing, and argue that requiring that data analyses are machine-readable can reduce the costs associated with replication.

We contribute to several literatures. The first is a burgeoning literature on the replication crisis, its causes, and prospective remedies (Loken and Gelman 2017; Hung and Fithian 2020; Jin et al. 2023). The second is a body of contemporary work in selective inference, in which Researcher decisions can be modelled as statistical features of a hypothesis testing problem (Fithian et al. 2017; Kuchibhotla et al. 2022; Goeman and Solari 2023; Andrews et al. 2023). The third approach uses formal methods to study incentive-compatibility problems in research design

[*]PhD Candidate, Department of Political Science, University of California, Berkeley. I am grateful to Kirk Bansak, Tak-Huen Chau, Thad Dunning, Will Fithian, Sean Gailmard, and Andrew Little for helpful comments and feedback.

Corresponding author:
Adam Bouyamourn, adam.bouyamourn@berkeley.edu

(Rosenthal 1979; Berinsky et al. 2021; Bates et al. 2022, 2023; Wang et al. 2023; Andrews and Shapiro 2021; Frankel and Kasy 2022).

In Political Science, Gerber et al. (2001), Gerber and Malhotra (2008), and Owen and Li (2021) provide empirical evidence of the existence of publication bias within the discipline. Amongst others, the Metaketa initiative has sought to institutionalize third-party reproduction of results as part of the research process (Dunning et al. 2019a,b). Journal replication and reproduction policies, and potential strategies to encourage more widespread replication, are described in Brodeur et al. (2024).

While our model specifically describes misreporting, its morals generalize to other examples of poor academic behaviour, and can help to explain situations in which a discipline tolerates other suboptimal research practices. Our work also highlights the intellectual value of a large, diverse research community, and the provides microfoundations for a healthy culture of cumulative learning through adversarial review.

The Misreporting Game

Setup

Actors

- A **Researcher**, who decides whether or not to misreport the results of a paper.
- A **Replicator**, who chooses whether or not to conduct an *ex post* replication of that paper.
- A **Readership**, of size N, who wishes to gain information about the state of the world and take an action.

Order of actions

1. Nature chooses the state of the world:

$$\omega = H\{H_0 \text{ is False}\} + (1-H)\{H_0 \text{ is True}\}$$

   Where: $H \sim Bernoulli(\beta)$.[*] Nature also chooses the type of the Replicator: $\theta \sim Beta(\rho, 1)$, for $\rho > 0$.
2. The Researcher observes the state of the world, and chooses whether to report "H_0 is T" or "H_0 is F".[†]
3. The Replicator chooses whether to pay a cost κ to Replicate (R) the paper, thereby learning and publicizing the correct state of the world, or not ($\neg R$).
4. The Reader chooses whether to Reject (F) or not to Reject ($\neg F$) the null hypothesis.
5. Payoffs are realized.

The game is illustrated in Figure 1.

Type Space There are two types of Researcher: one who encounters the state of the world when the null hypothesis is True, and one who encounters the state of the world when the null hypothesis is False.

Each Replicator has a type $\theta \in [0, 1]$, which characterizes the social connectedness of the Replicator to the Researcher. An exogenous parameter $\rho > 0$ characterizes the overall social connectedness of a given discipline: when ρ is high, the expected degree of social connectedness in a discipline is high.[‡]

The Researcher's Payoffs We say that a paper is *debunked* if the Replicator chooses to Replicate when the Researcher has misreported the state of the world. The Researcher earns -1 if their paper is debunked by the replicator (an injury to the Researcher's reputation), and 1 if it is either not debunked or not Replicated.[§] In addition, the Researcher earns ν (a *novelty* premium) if they report that the null is false, and the paper is not debunked or not Replicated by the Replicator (Chopra et al. 2023).

The Replicator's Payoffs A Replicator of type θ earns a reward when the Researcher's work is published. Likewise, the Replicator loses θ when the Researcher's work is debunked. This is essentially an *altruism parameter* (Rashevsky 1949; Rapoport 1956; Brew 1973), which captures the degree of preference alignment between the Replicator and the Researcher. Social connectedness between academics implies a degree of 'linked fate', or investedness in each other's careers. Close colleagues may wish to see each other succeed; likewise they may not wish to see each other suffer reputational harm. Perversely, individuals within a research subcommunity may fear for the reputation of their subcommunity if results within it are systematically debunked; this could sustain a dynamic in which they mutually decide not to examine each other's work too closely.

The Replicator earns $\chi \in (0, 1)$ if they *confirm* an existing finding; that is, if they choose to Replicate, and the Researcher did not misreport the state of the world. They earn 1 if they publish a debunking replication, which ensures that there is a "gotcha" premium earned for publishing a replication that contradicts an existing finding (Maniadis et al. 2017; Berinsky et al. 2021). The Replicator pays a fixed *cost* $\kappa > 0$ if they choose to Replicate. We can also interpret these payoffs as rewards to the Replicator for learning the state of the world: perhaps an external professional reward, or an internal 'warm glow' reward to the Replicator for advancing science. Hence a Replicator faces a conflict between their connectedness to the Researcher, and the professional and intellectual rewards from accurately replicating a paper.

We assume that a Replicator earns θ if they choose not to conduct a replication. We can interpret this as follows: if a Replicator is not aligned with the Researcher, and chooses not to Replicate, they forego both professional and intellectual rewards from learning the state of the world. Their decision to take the more costly action of Replicating

[*] Heuristically, β can be interpreted either as the (exogenous) statistical power of the Researcher's design, or of the 'difficulty' of the problem (Eberhardt 2010).

[†] To simplify notation, we refer to failure to reject the null in a given study as equivalent to the assertion by the researcher that "H_0 is T". The concerned reader may substitute the phrase "H_0 cannot be rejected" throughout.

[‡] Though we do not model this explicitly, we could think of this as the *vertex connectivity* of a graph that describes the discipline's social and professional network (e.g. Diestel 2017). Modelling networks of advising, co-authorship, or publication could provide us with a heuristic sense of how socially connected a given research subcommunity is.

[§] We abstract away from journal publication decisions in this game. This does not necessarily entail unbiased journal publication decisions, however: we can interpret the results stated below as conditional on journal publication decisions.

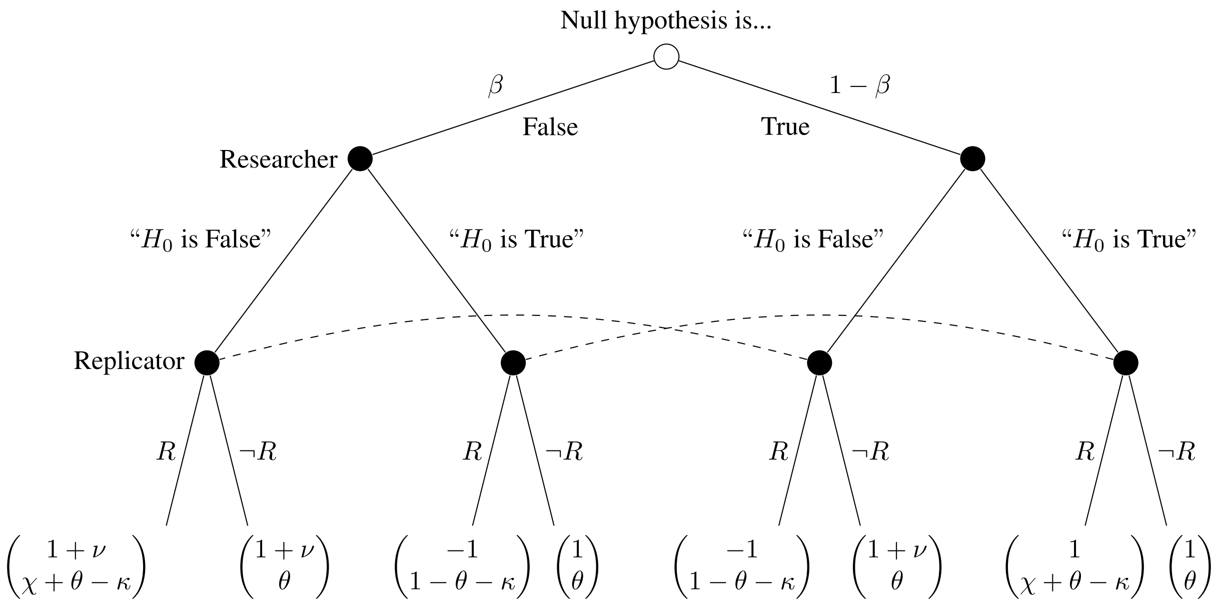

Figure 1. The Misreporting Game: The Researcher and Replicator's Subgame

a paper depends on its intrinsic and extrinsic benefits, social alignment with the paper's author, and the cost of Replicating the paper.

The Readership's Payoffs The Readership is composed of N readers who each earn 1 if they Reject the null hypothesis when and only when the null hypothesis is False, and -1 otherwise. Their beliefs about the state of the world are updated by Bayes' rule, and determine whether or not they choose to Reject the null.

Table 1. Model Parameters

Notation	Description
ρ	Connectedness of Researcher and Replicator pool
θ	Realized social connectedness
β	Probability that the Null is False
χ	Reward for publishing a confirmatory replication
ν	Reward for publishing a finding that rejects the Null
κ	Cost of conducting replication
γ	Probability of a random audit
λ	Probability that Replication reveals state

Equilibrium Concept

The relevant equilibrium concept is Perfect Bayesian Equilibrium. The separating equilibrium of interest is one in which the Researcher matches the state of the world to their report. The pooling equilibrium of interest is where the Researcher misreports the state of the world when the null hypothesis is in fact True, and accurately reports the state of the world when the null is False. We are interested in characterizing the pooling equilibrium in which both types of the Researcher report that the null is False. We do not study equilibrium in mixed strategies, since the morals are similar to those found in the pure strategy equilibrium. In what follows, we characterize the conditions that sustain an equilibrium in misreporting.

Solutions

To simplify notation, we write the states of the world as $\omega \in \Omega = \{T, F\}$, and we describe a message function $m : \Omega \to$ "Ω". We therefore write, e.g. $m(T) =$ "F" to describe the event that the researcher reports that "H_0 is F" when H_0 is in fact True.

We restrict our attention to parameter values for which 'interesting' equilibrium behaviour occurs.

Assumption 1. $\nu > 0$, $1 > \chi > \kappa > 0$, *and* $\beta < \frac{1}{2}$.

This ensures that the novelty premium is non-zero, which means that there is some positive inducement for the Reader to misreport results rather than accurately reporting the state of the world. Second, we have that the reward for debunking a study is greater than the reward for a confirmatory replication $(1 > \chi)$. Third, we have that $\chi > \kappa$, so that the cost of conducting a replication does not prevent the occurrence of replications in equilibrium. It should be noted that if κ is sufficiently large, no replications will ever occur. Finally, we suppose that $\beta < \frac{1}{2}$, or, that true positive results occur less frequently than true negative results.

First note that, when the null is in fact False, and the novelty premium is non-zero, the Researcher has no incentive to misreport their results.

Lemma 1. *Suppose* $\nu > 0$. *Then:*

$$U_{Researcher}(\text{"}H_0 \textit{ is } F\text{"}|H_0 \textit{ is } F) > U_{Res}(\text{"}H_0 \textit{ is } T\text{"}|H_0 \textit{ is } F)$$

This means that there is no pooling equilibrium in which the Researcher misreports that the null is True, when it is in fact False. We are interested in the pooling equilibrium in which misreporting occurs, so it is sufficient just to consider the Researcher's incentives to misreport in the case where the null is in fact True.

Replicator's Beliefs Let β^{post} be the Replicator's posterior belief that the null is False, given the Researcher's report. By

Bayes' rule, have that:

$$\beta^{post} = \frac{\beta}{\beta + (1-\beta)\mathbb{I}\{m(F) = \text{“}F\text{”} \wedge m(T) = \text{“}F\text{”}\}}$$

Where $\mathbb{I}\{m(F) = \text{“}F\text{”} \wedge m(T) = \text{“}F\text{”}\}$ is an indicator that denotes the event that the Researcher chooses the messaging strategy in which they always report that the null is False, irrespective of the actual state of the world. In a misreporting equilibrium, the types of the Researcher *pool*, and both report that the null is False. Then the Researcher's signal is entirely uninformative, and the Replicator believes that the null is False with $\beta^{post} = \beta$.

If the Researcher *separates*, and reports that the null is False when and only when the null is False, then the Researcher's signal is completely informative, and the Replicator believes that the null is True with certainty, given that the Researcher tells them so.

Replicator's Incentive-Compatibility Constraint We next consider when the Replicator will prefer to Replicate the paper. We show that there is a degree of social connectedness between Replicator and Researcher above which the Replicator will choose not to conduct a Replication.

Lemma 2. Replicator's I-C Constraint. *There is a cutpoint type $\tilde{\theta}(\beta, \chi, \kappa)$ such that, in any separating equilibrium, all types $\theta \leq \tilde{\theta}(\beta, \chi, \kappa)$ Replicate, and all types $\theta > \tilde{\theta}(\beta, \chi, \kappa)$ do not Replicate.*

That is, all types $\theta < \tilde{\theta}$ are sufficiently unaligned with the Researcher that they would prefer to Replicate and debunk the paper, rather than benefiting from the Researcher's success.

Researcher's Incentive-Compatibility Constraint The Researcher prefers to misreport when the expected payoff to doing so is positive. Their payoff depends on whether the Replicator will replicate their work.

Since $\theta \sim Beta(\rho, 1)$, the event $\{\theta \leq \tilde{\theta}(\beta, \chi, \kappa)\}$ occurs with probability that is known to the Researcher. We have:

$$Pr\left(\theta \leq \tilde{\theta}(\beta, \chi, \kappa)\right) = \left[\frac{\beta\chi - \kappa}{2(1-\beta)} + \frac{1}{2}\right]^{\rho}$$

Next, we characterize the circumstances under which the Researcher prefers to pool; that is, to report that the null is False in any state of the world. We previously established, in Lemma 1, that there is no pooling equilibrium in which the Researcher misreports that the null is True when it is in fact False.

For the Researcher to prefer to misreport, we have:

Lemma 3. Researcher's I-C Constraint. *There is a threshold $\tilde{\rho}(\nu, \beta, \chi, \kappa)$ such that, if $\rho > \tilde{\rho}(\nu, \beta, \chi, \kappa)$, the Researcher prefers to misreport when the null hypothesis is True.*

In words, there is a degree of social connectivity above which the Researcher prefers to misreport: this entails that the degree of preference alignment between the Researcher and Replicator pool is sufficiently large that the Researcher will prefer to misreport and 'take their chances' that the Replicator will be sufficiently socially-aligned to prefer not to debunk the paper.

Readership's Beliefs We suppose that a Replication is fully informative, so that the Readership can always take the correct action when a Replication occurs. In this case, their posterior beliefs match the state of the world. When a Replication does not occur, the Readership knows that this is because the Replicator is sufficiently socially aligned with the Researcher; but since the Replicator does not know the state of the world, this does not convey information to the Readership about the state of the world. In this case, the Readership's beliefs are equal to their prior beliefs.

Readership's Strategy When the Replicator chooses not to Replicate, the Readership's posterior belief that the null is False is β, the common-knowledge prior probability that the null is False. In this case, since $\beta < \frac{1}{2}$, the Readership chooses to not to Reject the null hypothesis., since doing so minimizes their misclassification error with respect to the state of the world.

Lemma 4. Readership's Strategy.

$$\beta < \frac{1}{2} \implies \mathbb{E}[U_{Reader}(\neg F|\neg R)] > \mathbb{E}[U_{Reader}(F|\neg R)]$$

Equilibrium

Proposition 1. Collusive and Adversarial Replication.

Suppose Assumption 1 holds. Then:

1. ***Collusive Equilibrium**. If Replicator is sufficiently socially connected to the Researcher, that is, when $\rho \geq \tilde{\rho}(\nu, \beta, \chi, \kappa)$, there is a pooling PBE in which the Researcher misreports their results, and a cutpoint type of the Replicator $\tilde{\theta}(\beta, \chi, \kappa)$ such that all Replicators of type $\theta > \tilde{\theta}(\beta, \chi, \kappa)$ choose not to Replicate. If no Replication occurs, the Readership chooses not to reject the null hypothesis; otherwise they accurately match their action to the state of the world.*
2. ***Adversarial Equilibrium.** Otherwise, there is a separating PBE in which the Researcher reports their results accurately, all types of the Replicator replicate, and the Readership accurately matches their belief to the state.*

When the pool of Replicators is sufficiently socially connected on average, there is an equilibrium in *collusion*. That is, the Researcher faces a Replicator pool that is sufficiently highly aligned that they prefer to misreport. We call this the Collusive Equilibrium, because it is one in which Researchers and Replicators effectively collude to ensure that bad research does not get debunked.

When socially connected Replicators opt out of Replication, this encourages misreporting, because it enables authors to evade the possibility of seeing their papers debunked *ex post*. By contrast, when the Replicator pool is sufficiently heterogeneous, the risk of *ex post* review is sufficiently large that Researchers can be dissuaded from misreporting results.

Social Welfare

Researcher and Replicator incentives matter because research has social value. Research is a social enterprise that aims to improve public knowledge about the world:

misreporting causes harm to the public. We give this a concrete interpretation in terms of benefits to the Readership.

The False Discovery Rate (FDR) The False Discovery Rate (FDR) is a concept from the multiple testing and selective inference literature. It describes the number of hypothesis tests in a collection that falsely reject the null when it is in fact True (Seeger 1968; Benjamini and Hochberg 1995, 2000; Storey 2003; Efron 2010). In our model, social welfare has an intuitive interpretation: it is a function of the proportion of published studies that misreport results.

Because we consider a somewhat simplified hypothesis testing problem, we define the (field-wise, Bayesian) FDR in the following way.

Definition 1. Field-wise, Bayesian False Discovery Rate.

$$FDR \equiv Pr(H_0 \textit{ is True} \,|\, \text{``}H_0 \textit{ is False''}, \neg R)$$

That is, the probability that the Researcher misreports, *and* their report is not debunked by the Replicator.

The first thing to note is that this is a Bayesian notion of false discovery, given both the Researcher's and the Replicator's report. This is closest to the definition provided in Bates et al. (2023). The second is that False Discovery Rate here is *field-wise*: it represents the FDR given the actions of multiple actors. Hence it is a measure of *cumulative* failures of learning. In what follows we refer to this quantity as the FDR, though noting the differences in context and setting from other uses of this concept.

We study the equilibrium in misreporting from proposition 1, that is, when $\rho \geq \tilde{\rho}(\nu, \beta, \chi, \kappa)$. First, we can write the FDR as a function of model parameters, as follows:

Lemma 5. False Discovery Rate.

$$FDR(\beta, \chi, \kappa, \rho) = (1-\beta)\left[1 - \tilde{\theta}(q, \chi, \kappa)^{\rho}\right]$$

Intuitively, this is a product of two terms: the probability that the null is true, since misreporting can only occur when the null is true; and the probability that a Replicator is sufficiently socially aligned with the Researcher to choose not to Replicate. Misreporting occurs and is not debunked when *both* of these events occur.

We also note that the main loss of welfare in the collusive equilibrium occurs because the Readership no longer trusts the conclusions of research. No belief updating occurs as a result of publication, and the Readership maintains their prior beliefs. Collusion reduces trust in the research process.

Expressing Social Welfare In Terms of the FDR We now study social welfare across the two equilibria of The Misreporting Game.

Proposition 2. Social Welfare Loss Due to Collusion.

i. *There is a minimum readership size N_0 such that $N > N_0$ implies there is net social welfare loss from moving to the collusive equilibrium from the adversarial equilibrium.*
ii. *When $N > N_0$, the magnitude of the social welfare loss due to collusion is increasing in the False Discovery Rate.*

Intuitively social welfare is maximized in the separating equilibrium, in which the Researcher reports the state of the world truthfully, and all types of Replicator replicate.

Our first observation is that, in order for the collusive equilibrium to be harmful, there must be a sufficiently large readership. In other words, research has to *matter*.

The second observation is that the False Discovery Rate is a proxy for social welfare loss in this model. This is intuitive, but highlights the social dimension of the FDR as a measure of the quality of a research practice. We now turn to assessing its comparative statics.

Comparative Statics

Since social welfare is monotonically decreasing in the FDR, the main morals of our model can be found by analysing the FDR.

Proposition 3. Comparative Statics.

i. *The FDR is increasing in ρ (the degree of social connectedness), and κ (the cost of replication).*
ii. *The FDR is decreasing in χ (the incentive to confirm existing studies) and β (the plausibility of the alternative hypothesis or the power of the statistical procedure).*
iii. *The threshold value of ρ needed to sustain the Collusive Equilibrium is decreasing in ν (the return to novelty).*

Social Connectedness

$$\frac{\partial FDR(\beta, \chi, \kappa, \rho)}{\partial \rho} > 0$$

Our first result is that the FDR increases as social connectedness between Researchers and Replicators increases. When social connectedness is high, Replicators can be discouraged from conducting Replications. Improving the heterogeneity of the academic discipline, and hence, of the Replicator pool, improves social welfare. This helps to ensure that replication is *adversarial*, rather than *collusive*.

Academic fields in which there is a more heterogeneous pool of peers, and weaker connections between peers, are likely to do a better job of advancing knowledge. Further, this suggests that *larger* academic fields are less likely to face systematic problems of misreporting: it is harder to sustain a cultural equilibrium in which egregious errors are not overturned. This provides a incentive-based argument for the common claim that a variety of intellectual, and personal backgrounds is important for encouraging fraud detection, and for preventing the growth of collective blindspots.

Incentive To Confirm Existing Studies

$$\frac{\partial FDR(\beta, \chi, \kappa, \rho)}{\partial \chi} < 0$$

We also highlight that incentives to perform confirmatory replications are important. Replicators are more likely to *enter* when there are strong career incentives to replicate papers.

Costs of Replication

$$\frac{\partial FDR(\beta, \chi, \kappa, \rho)}{\partial \kappa} > 0$$

High costs, perhaps due to poor coding practices and data availability, discourage replications. This is intuitive, in that the costs of replication naturally affect how desirable it is for Replicators to conduct them. Nonetheless, we can interpret this somewhat more broadly. First, suppose that the cost of replication is determined by the protocols followed by journals in publishing original research. Journals commonly require standards for the reproducibility of papers, where reproducibility refers to the ease with which the original analysis and its exact conclusions are recreated by a third-party analyst. Requiring more stringent standards for reproducibility should lower the costs to third-parties of replicating papers, which in turn decreases the FDR and improves social welfare.

Plausibility of the Alternative

$$\frac{\partial FDR(\beta, \chi, \kappa, \rho)}{\partial \beta} < 0$$

Interpreting β as the probability that the null is in fact False, the first-order effect on the FDR is that making a test easier to reject reduces the need to misreport results in the first place. If the null were always False, there would be no incentive to misreport.

There is a subtler effect on Replicator incentives, however, where:

$$\frac{\partial \tilde{\theta}(\beta, \chi, \kappa)}{\partial \beta} < 0$$

Inspection of the Replicator's Incentive-Compatibility constraint highlights that the Replicator earns a premium from debunking a paper equal to $\frac{1-\chi}{2}$. It is only possible to earn this premium, however, when the null is in fact True, since this is the only state of the world in which misreporting occurs. Hence, when the null is False more often, fewer types of the Replicator find it profitable to Replicate. This means that a smaller degree of preference alignment between Researcher and Replicator is needed for the Replicator to choose not to Replicate.

Relative Rewards for Novelty The returns to novelty do not directly affect the FDR, since they do not factor into the Replicator's decision whether or not to Replicate the article. However, the novelty premium does reduce the required degree of social connectedness needed to sustain a pooling equilibrium to begin with. We have:

$$\frac{\partial \tilde{\rho}(\nu, \beta, \chi, \kappa)}{\partial \nu} < 0$$

It is easier to sustain an equilibrium in misreporting when the rewards for novelty are relatively higher: the Researcher's best response is less sensitive to the probability of detection. This is intuitive, since there is a greater return to successful fabrication by the Researcher.

Extension: Random Audits

Now we suppose that, with probability γ, the paper is assigned to an external, unaligned Replicator, who must Replicate the paper. This is independent of the probability that a randomly-drawn replicator chooses to replicate the paper.

In this setting, the probability of Replication becomes:

$$Pr(R|\gamma, \theta) = \gamma + (1-\gamma)Pr(\theta \leq \tilde{\theta})$$

We state the conclusions of this game below, reserving analysis to Appendix A.2.

Proposition 4. Random Audits.

Relative to the Misreporting game:

i. *The threshold value $\tilde{\rho}(\gamma)$ is larger, meaning that a higher degree of social connectedness, and a minimum novelty premium $\tilde{\nu}(\gamma)$ required to sustain the collusive equilibrium.*
ii. *The FDR is smaller for any nonzero value of γ.*
iii. *The FDR is decreasing, and hence social welfare is increasing, in γ.*

The moral is that random auditing can reduce the possibility of misreporting induced by the selective entrance of Replicators. To counterbalance the effect of random auditing, a higher level of discipline-wide social connectivity is required to sustain the collusive equilibrium.

Extension: Noisy Replication

We next consider the case that, instead of revealing the state with certainty, Replication successfully reveals the state with probability λ. This generalizes the Misreporting Game, when replication is perfectly informative, and $\lambda = 1$.

Proposition 5.

There exists a threshold value λ^*, such that:

i. *When $\lambda \leq \lambda^*$ entails that Replication is uninformative, the Readership follows its prior belief in all circumstances, there is no social welfare loss from collusion as the research process does not inform the Readership's action.*
ii. *When $\lambda > \lambda^*$, Replication informs the Readership's action, and there is positive social welfare loss from collusion.*
iii. *The equilibrium described in Proposition 1 characterizes the case where $\lambda = 1$.*

We provide an analysis of this case in Appendix A.3. Essentially, social welfare loss due to collusion can be understood as proportional to the information that the Readership would have received had collusion not taken place. The more informative Replication is, the more information is lost to collusion, and hence, the greater the social welfare loss.

Discussion

Collusive and Adversarial Replication

In the Misreporting Game, the degree of preference alignment between Researchers and Replicators is harmful to social welfare: when Replicators and Researchers have a high degree of preference alignment, it is easier to sustain an equilibrium in Misreporting. While this most obviously applies to the case of outright fraud, an analogous argument applies to any kind of undesirable academic practice that could be identified in an *ex post* Replication study. For instance, selecting on the significance of a result, p-hacking, site selection, or model hacking could all be sustained by similar incentive-compatibility problems as described in the above setting (Button et al. 2013; Head et al. 2015; Allcott 2015; Deaton and Cartwright 2018; Kuchibhotla et al. 2022). Further, academic fraud need not be intentional: researchers may simply interpret ambiguous signals in ways that self-servingly cohere with their prior beliefs (Babcock and Loewenstein 1997).

A high degree of social alignment between Researchers and Replicators can therefore help to sustain cultural norms that are permissive of suboptimal research practices. The main mechanism by which this harms social welfare is a reduction in trust in the research process: in the absence of solid ex post replication, readers may be less willing to endorse the conclusions of a study.

Considering the determinants of preference alignment between members of a research community provides us with a strong argument in favour of increasing the diversity of academic fields, broadly construed. The extent to which a field is numerically large, comprised of members of a wide array of institutions, with intellectually diverse memberships, among whom open debate and the challenging of long-held assumptions is encouraged, can have a tangible effect on the reliability of research, and hence, its social value.

One approach to inducing good behaviour by Researchers is institutionalizing professional benefits to having their results successfully replicated. Tenure review decisions represent an important professional milestone and one of the last formal institutional opportunities for rewards to be conditioned on research quality: assigning greater weight to the replicability, or external validity of a researcher's output at this stage could plausibly induce better research practices.

Other strategic dynamics that we have not modelled may also be relevant to equilibrium replication behaviour. An alternative form of collusion could involve strategic misrepresentation of the results of replications by Replicators: in practice, readers may face the problem that meta-analyses or replications are *also* somewhat misleading about the state of the world. Finally, our results consider a worst-case version of Researcher and Replicator incentives: ideally, both Researchers and Replicators are motivated in all circumstances by an excessive zeal to accurately report the state of the world.

Random Auditing: Fire-Alarm versus Police-Patrol Replication

McCubbins and Schwartz (1984) famously contrasted *fire-alarm oversight*, a decentralized system in which third parties are incentivized to intervene on issues of importance, with *police-patrol oversight*, in which centralized auditing of legislative outputs occurs.

Currently, replication efforts in political science are decentralized, and rely on the efforts of individuals: we could broadly characterize the status quo as *fire-alarm replication*. We have argued that this is compatible with an equilibrium in which third parties do not replicate papers they would otherwise replicate due to social connections, in addition to career concerns, poor coding practices, and journal biases against publishing replications. Moving towards police-patrol replication, as discussed above, should improve FDR control. Further, this argues in favour of increasing the discipline-wide allocation of resources towards encouraging centralized efforts at Replication.

Automated Code Tests

In technology companies engaged in software engineering, a standard part of the workflow is to use unit and acceptance tests to automatically assess the interoperability of code written by individual programmers (Cheon and Leavens 2002; Runeson 2006; Winters et al. 2020). Some journals require that analyses are reproducible by humans, but it is possible to automate this procedure. Requiring hygienic coding practices that can be machine-checked by online tools prior to publication is a feasible way to improve the reproducibility of articles. This could help to reduce the cost of replicating articles, which we have identified as a determinant of the FDR.

Reducing the Statistical Significance Threshold

The results in Proposition 3 bear relevantly on a recent proposal to reduce the threshold for significance from $p \leq .05$ to $p \leq .005$ (Benjamin et al. 2017). The logic of the above argument highlights that doing so could actually *increase* misreporting: requiring a higher standard of evidence may induce Researchers to misreport more often. A higher rate of misreporting also encourages more Replicators to conduct replications; but Replicator entrance is not perfectly elastic with respect to β, since it also depends on the payoff earned from conducting a confirmatory replication, so, the FDR increases overall. Further, reducing the statistical significance threshold could increase the cost of replication, which would further increase the FDR. Since this is not the primary focus of this article, however, we assert simply that it is insufficient to consider the statistical implications of adjusting the conventional threshold for significance alone: how it affects the equilibrium behaviour of researchers, reviewers and replicators must also be considered.

Conclusions

Designing socially-optimal incentive schemes for replication is an important task as social science incorporates lessons from the Replication Crisis, and from recent literature in selective inference. We have highlighted the role of social connectedness, and shown that it can have a malign impact on academic standards in the publication process if it discourages replications. We advocate in favour of greater allocation of resources towards auditing of papers by both institutions and journals. Moving towards police-patrol replication should improve the reliability and social value of research. Lowering the statistical significance threshold may not have the intended effect on researcher incentives. Above all, the problem of ensuring that research is high-quality and replicable is a problem of incentive-compatibility: proposals to improve research and replication practices must pay close attention to the incentives of members of the research community.

References

Ahmadov AK (2014) Oil, democracy, and context: A meta-analysis. *Comparative Political Studies* 47(9): 1238–1267. DOI:10.1177/0010414013495358. URL https://doi.org/10.1177/0010414013495358.

Allcott H (2015) Site Selection Bias in Program Evaluation *. *The Quarterly Journal of Economics* 130(3): 1117–1165. DOI:10.1093/qje/qjv015. URL https://doi.org/10.1093/qje/qjv015.

Anastasopoulos N and Asteriou D (2019) Optimal dynamic auditing based on game theory. *Operational Research* 21(3): 1887–1912. DOI:10.1007/s12351-019-00491-3. URL http://dx.doi.org/10.1007/s12351-019-00491-3.

Andrews I and Kasy M (2019) Identification of and correction for publication bias. *American Economic Review* 109(8): 2766–94. DOI:10.1257/aer.20180310. URL https://www.aeaweb.org/articles?id=10.1257/aer.20180310.

Andrews I, Kitagawa T and McCloskey A (2023) Inference on Winners*. *The Quarterly Journal of Economics* 139(1): 305–358. DOI:10.1093/qje/qjad043. URL https://doi.org/10.1093/qje/qjad043.

Andrews I and Shapiro JM (2021) A model of scientific communication. *Econometrica* 89(5): 2117–2142. DOI:https://doi.org/10.3982/ECTA18155. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA18155.

Babcock L and Loewenstein G (1997) Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic Perspectives* 11(1): 109–126. DOI:10.1257/jep.11.1.109. URL https://www.aeaweb.org/articles?id=10.1257/jep.11.1.109.

Bates S, Jordan MI, Sklar M and Soloff JA (2022) Principal-agent hypothesis testing.

Bates S, Jordan MI, Sklar M and Soloff JA (2023) Incentive-theoretic bayesian inference for collaborative science.

Bednar J and Page S (2007) Can game(s) theory explain culture?: The emergence of cultural behavior within multiple games. *Rationality and Society* 19(1): 65–97. DOI:10.1177/1043463107075108. URL https://doi.org/10.1177/1043463107075108.

Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, Bollen KA, Brembs B, Brown L, Camerer C, Cesarini D, Chambers CD, Clyde M, Cook TD, De Boeck P, Dienes Z, Dreber A, Easwaran K, Efferson C, Fehr E, Fidler F, Field AP, Forster M, George EI, Gonzalez R, Goodman S, Green E, Green DP, Greenwald AG, Hadfield JD, Hedges LV, Held L, Hua Ho T, Hoijtink H, Hruschka DJ, Imai K, Imbens G, Ioannidis JPA, Jeon M, Jones JH, Kirchler M, Laibson D, List J, Little R, Lupia A, Machery E, Maxwell SE, McCarthy M, Moore DA, Morgan SL, Munafó M, Nakagawa S, Nyhan B, Parker TH, Pericchi L, Perugini M, Rouder J, Rousseau J, Savalei V, Schönbrodt FD, Sellke T, Sinclair B, Tingley D, Van Zandt T, Vazire S, Watts DJ, Winship C, Wolpert RL, Xie Y, Young C, Zinman J and Johnson VE (2017) Redefine statistical significance. *Nature Human Behaviour* 2(1): 6–10. DOI:10.1038/s41562-017-0189-z. URL http://dx.doi.org/10.1038/s41562-017-0189-z.

Benjamini Y and Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1): 289–300. URL http://www.jstor.org/stable/2346101.

Benjamini Y and Hochberg Y (2000) On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics* 25(1): 60–83. DOI:10.3102/10769986025001060. URL https://doi.org/10.3102/10769986025001060.

Berinsky AJ, Druckman JN and Yamamoto T (2021) Publication biases in replication studies. *Political Analysis* 29(3): 370–384. DOI:10.1017/pan.2020.34.

Brew JS (1973) An altruism parameter for prisoner's dilemma. *The Journal of Conflict Resolution* 17(2): 351–367. URL http://www.jstor.org/stable/173228.

Brodeur A, Esterling K, Ankel-Peters J, Bueno NS, Desposato S, Dreber A, Genovese F, Green DP, Hepplewhite M, de la Guardia FH, Johannesson M, Kotsadam A, Miguel E, Velez YR and Young L (2024) Promoting reproducibility and replicability in political science. *Research & Politics* 11(1): 20531680241233439. DOI:10.1177/20531680241233439. URL https://doi.org/10.1177/20531680241233439.

Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ and Munafò MR (2013) Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience* 14(5): 365–376. DOI:10.1038/nrn3475. URL http://dx.doi.org/10.1038/nrn3475.

Camerer CF, Dreber A, Holzmeister F, Ho TH, Huber J, Johannesson M, Kirchler M, Nave G, Nosek BA, Pfeiffer T, Altmejd A, Buttrick N, Chan T, Chen Y, Forsell E, Gampa A, Heikensten E, Hummer L, Imai T, Isaksson S, Manfredi D, Rose J, Wagenmakers EJ and Wu H (2018) Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nature Human Behaviour* 2(9): 637–644. DOI:10.1038/s41562-018-0399-z. URL http://dx.doi.org/10.1038/s41562-018-0399-z.

Cheon Y and Leavens GT (2002) A simple and practical approach to unit testing: The jml and junit way. In: *Proceedings of the 16th European Conference on Object-Oriented Programming*, ECOOP '02. Berlin, Heidelberg: Springer-Verlag. ISBN 3540437592, p. 231–255.

Chopra F, Haaland I, Roth C and Stegmann A (2023) The Null Result Penalty. *The Economic Journal* : uead060DOI:10.1093/ej/uead060. URL https://doi.org/10.1093/ej/uead060.

Collaboration OS (2015) Estimating the reproducibility of psychological science. *Science* 349(6251): aac4716. DOI:10.1126/science.aac4716. URL https://www.science.org/doi/abs/10.1126/science.aac4716.

Davis AM, Flicker B, Hyndman K, Katok E, Keppler SM, Leider S, Long X and Tong JD (2023) A replication study of operations management experiments in management science. *SSRN Electronic Journal* URL https://api.semanticscholar.org/CorpusID:249840611.

Deaton A and Cartwright N (2018) Understanding and misunderstanding randomized controlled trials. *Social Science Medicine* 210: 2–21. DOI:https://doi.org/10.1016/j.socscimed.2017.12.005. URL https://www.sciencedirect.com/science/article/pii/S0277953617307359. Randomized Controlled Trials and Evidence-based Policy: A Multidisciplinary Dialogue.

Diestel R (2017) *Graph Theory*. 5th edition. Springer Publishing Company, Incorporated. ISBN 3662536218.

Dunning T, Grossman G, Humphreys M, Hyde SD, McIntosh C and (Eds) GN (2019a) *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*. Cambridge Studies in Comparative Politics. Cambridge University Press.

Dunning T, Grossman G, Humphreys M, Hyde SD, McIntosh C, Nellis G, Adida CL, Arias E, Bicalho C, Boas TC, Buntaine MT, Chauchard S, Chowdhury A, Gottlieb J, Hidalgo FD, Holmlund M, Jablonski R, Kramon E, Larreguy H, Lierl M, Marshall J, McClendon G, Melo MA, Nielson DL, Pickering PM, Platas MR, Querubín P, Raffler P and Sircar N (2019b) Voter information campaigns and political accountability: Cumulative findings from a preregistered meta-analysis of coordinated trials. *Science Advances* 5(7): eaaw2612. DOI:10.1126/sciadv.aaw2612. URL https://www.science.org/doi/abs/10.1126/sciadv.aaw2612.

Eberhardt F (2010) Causal discovery as a game. In: Guyon I, Janzing D and Schölkopf B (eds.) *Proceedings of Workshop on Causality: Objectives and Assessment at NIPS 2008*, *Proceedings of Machine Learning Research*, volume 6. Whistler, Canada: PMLR, pp. 87–96. URL https://proceedings.mlr.press/v6/eberhardt10a.html.

Efron B (2010) *Large-Scale Hypothesis Testing*. Institute of Mathematical Statistics Monographs. Cambridge University Press, p. 15–29. DOI:10.1017/CBO9780511761362.003.

Esarey J and Wu A (2016) Measuring the effects of publication bias in political science. *Research & Politics* 3(3): 2053168016665856. DOI:10.1177/2053168016665856. URL https://doi.org/10.1177/2053168016665856.

Fithian W, Sun D and Taylor J (2017) Optimal inference after model selection.

Fourcade M (2009) *Economists and Societies: Discipline and Profession in the United States, Britain, and France, 1890s to 1990s*. Princeton Studies in Cultural Sociology. Princeton University Press. ISBN 9781400833139. URL https://books.google.com/books?id=lggyMnpVf5kC.

Francis G (2013) Replication, statistical consistency, and publication bias. *Journal of Mathematical Psychology* 57(5): 153–169. DOI:https://doi.org/10.1016/j.jmp.2013.02.003. URL https://www.sciencedirect.com/science/article/pii/S002224961300014X. Special Issue: A Discussion of Publication Bias and the Test for Excess Significance.

Franco A, Malhotra N and Simonovits G (2014) Publication bias in the social sciences: Unlocking the file drawer. *Science* 345(6203): 1502–1505. DOI:10.1126/science.1255484. URL https://www.science.org/doi/abs/10.1126/science.1255484.

Frankel A and Kasy M (2022) Which findings should be published? *American Economic Journal: Microeconomics* 14(1): 1–38. DOI:10.1257/mic.20190133. URL https://www.aeaweb.org/articles?id=10.1257/mic.20190133.

Gelman A and Loken E (2013) The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time URL http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf.

Gerber A and Malhotra N (2008) Do statistical reporting standards affect what is published? publication bias in two leading political science journals. *Quarterly Journal of Political Science* 3. DOI:10.1561/100.00008024.

Gerber AS, Green DP and Nickerson D (2001) Testing for publication bias in political science. *Political Analysis* 9(4): 385–392. URL http://www.jstor.org/stable/25791658.

Goeman J and Solari A (2023) On selecting and conditioning in multiple testing and selective inference.

Goodman SN, Fanelli D and Ioannidis JPA (2016) What does research reproducibility mean? *Science Translational Medicine* 8(341): 341ps12–341ps12. DOI:10.1126/scitranslmed.aaf5027. URL https://www.science.org/doi/abs/10.1126/scitranslmed.aaf5027.

Head ML, Holman L, Lanfear R, Kahn AT and Jennions MD (2015) The extent and consequences of p-hacking in science. *PLOS Biology* 13(3): e1002106. DOI:10.1371/journal.pbio.1002106. URL http://dx.doi.org/10.1371/journal.pbio.1002106.

Howe PDL and Perfors A (2018) An argument for how (and why) to incentivise replication. *Behavioral and Brain Sciences* 41: e135. DOI:10.1017/S0140525X18000705.

Hung K and Fithian W (2020) Statistical methods for replicability assessment. *The Annals of Applied Statistics* 14(3). DOI:10.1214/20-aoas1336. URL https://doi.org/10.1214%2F20-aoas1336.

Ioannidis JPA (2005) Why most published research findings are false. *PLOS Medicine* 2(8): null. DOI:10.1371/journal.pmed.0020124. URL https://doi.org/10.1371/journal.pmed.0020124.

Jin Y, Guo K and Rothenhausler D (2023) Diagnosing the role of observable distribution shift in scientific replications. URL https://api.semanticscholar.org/CorpusID:261530919.

Johnson VE, Payne RD, Wang T, Asher A and Mandal S (2017) On the reproducibility of psychological science. *Journal of the American Statistical Association* 112(517): 1–10. DOI:10.1080/01621459.2016.1240079. URL http://dx.doi.org/10.1080/01621459.2016.1240079.

Kuchibhotla AK, Kolassa JE and Kuffner TA (2022) Post-selection inference. *Annual Review of Statistics and Its Application* 9(1): 505–527. DOI:10.1146/annurev-statistics-100421-044639. URL https://doi.org/10.1146/annurev-statistics-100421-044639.

Leamer EE (1974) False models and post-data model construction. *Journal of the American Statistical Association* 69(345): 122–131. URL http://www.jstor.org/stable/2285510.

Lehmann E and Romano J (2005) *Testing Statistical Hypotheses*. Springer New York. ISBN 9780387276052. DOI:10.1007/0-387-27605-x. URL http://dx.doi.org/10.1007/

0-387-27605-X.

Lewis DK (1969) *Convention: A Philosophical Study*. Cambridge, MA, USA: Wiley-Blackwell.

Loken E and Gelman A (2017) Measurement error and the replication crisis. *Science* 355(6325): 584–585. DOI:10.1126/science.aal3618. URL https://www.science.org/doi/abs/10.1126/science.aal3618.

Lundwall RA (2019) Changing institutional incentives to foster sound scientific practices: One department. *Infant Behavior and Development* 55: 69–76. DOI: https://doi.org/10.1016/j.infbeh.2019.03.006. URL https://www.sciencedirect.com/science/article/pii/S0163638318300900.

Lupia A (2021) Practical and ethical reasons for pursuing a more open science. *PS: Political Science amp; Politics* 54(2): 301–304. DOI:10.1017/S1049096520000979.

Maniadis Z, Tufano F and List JA (2017) To replicate or not to replicate? exploring reproducibility in economics through the lens of a model and a pilot study. *The Economic Journal* 127(605): F209–F235. DOI:https://doi.org/10.1111/ecoj.12527. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ecoj.12527.

McCubbins MD and Schwartz T (1984) Congressional oversight overlooked: Police patrols versus fire alarms. *American Journal of Political Science* 28(1): 165–179. URL http://www.jstor.org/stable/2110792.

Miguel E (2021) Evidence on research transparency in economics. *Journal of Economic Perspectives* 35(3): 193–214. DOI: 10.1257/jep.35.3.193. URL https://www.aeaweb.org/articles?id=10.1257/jep.35.3.193.

Murray MK and Rice JW (1993) *Differential geometry and statistics*. Chapman & Hall/CRC Monographs on Statistics and Applied Probability. Philadelphia, PA: Chapman & Hall/CRC.

Owen E and Li Q (2021) The conditional nature of publication bias: a meta-regression analysis. *Political Science Research and Methods* 9(4): 867–877. DOI:10.1017/psrm.2020.15.

Patty JW and Penn EM (2020) Identity and information in organizations. *Journal of Political Institutions and Political Economy* 1(3): 379–416. DOI:10.1561/113.00000014. URL http://dx.doi.org/10.1561/113.00000014.

Rapoport A (1956) Some game-theoretical aspects of parasitism and symbiosis. *The Bulletin of Mathematical Biophysics* 18(1): 15–30. DOI:10.1007/bf02477841. URL http://dx.doi.org/10.1007/BF02477841.

Rashevsky N (1949) Mathematical biology of social behavior. *The Bulletin of Mathematical Biophysics* 11(2): 105–113. DOI: 10.1007/bf02477497. URL http://dx.doi.org/10.1007/BF02477497.

Rosenthal R (1979) The file drawer problem and tolerance for null results. *Psychological Bulletin* 86(3): 638–641. DOI:10.1037/0033-2909.86.3.638. URL http://dx.doi.org/10.1037/0033-2909.86.3.638.

Runeson P (2006) A survey of unit testing practices. *IEEE Software* 23(4): 22–29. DOI:10.1109/ms.2006.91. URL http://dx.doi.org/10.1109/MS.2006.91.

Seeger P (1968) A note on a method for the analysis of significances en masse. *Technometrics* 10(3): 586–593. DOI:10.1080/00401706.1968.10490605. URL https://www.tandfonline.com/doi/abs/10.1080/00401706.1968.10490605.

Slough T and Tyson SA (2024) Sign-congruence, external validity, and replication. *Political Analysis* URL https://api.semanticscholar.org/CorpusID:254283845.

Sterling TD, Rosenbaum WL and Weinkam JJ (1995) Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician* 49(1): 108–112. URL http://www.jstor.org/stable/2684823.

Stirling WC and Felin T (2013) Game theory, conditional preferences, and social influence. *PLoS ONE* 8(2): e56751. DOI:10.1371/journal.pone.0056751. URL http://dx.doi.org/10.1371/journal.pone.0056751.

Storey JD (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(3): 479–498. DOI:https://doi.org/10.1111/1467-9868.00346. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00346.

Storey JD (2003) The positive false discovery rate: a bayesian interpretation and the q-value. *The Annals of Statistics* 31(6). DOI:10.1214/aos/1074290335. URL https://doi.org/10.1214/aos/1074290335.

van Aert RCM and van Assen MALM (2017) Bayesian evaluation of effect size after replicating an original study. *PLOS ONE* 12(4): 1–23. DOI:10.1371/journal.pone.0175302. URL https://doi.org/10.1371/journal.pone.0175302.

Wang S, Bates S, Aronow PM and Jordan MI (2023) Operationalizing counterfactual metrics: Incentives, ranking, and information asymmetry.

Winters T, Manshreck T and Wright H (2020) *Software Engineering at Google: Lessons Learned from Programming Over Time*. O'Reilly Media. ISBN 9781492082743. URL https://books.google.com/books?id=WXTTDwAAQBAJ.

Wuttke A (2019) Why too many political science findings cannot be trusted and what we can do about it. *Politische Vierteljahresschrift* 60(1): 1–19.

Appendix

A.1 Proofs of Main Results

Proof of Lemma 1

Lemma 1. *Suppose* $\nu > 0$*. Then:*

$$\mathbb{E}_\theta[U_{Res}(\text{“}H_0 \textit{ is } F\text{”}|H_0 \textit{ is } F)] > \mathbb{E}_\theta[U_{Res}(\text{“}H_0 \textit{ is } T\text{”}|H_0 \textit{ is } F)]$$

Proof.

$$\begin{aligned}\min U_{Res}(\text{“}H_0 \text{ is F”}|H_0 \text{ is F}) &= 1+\nu \\ \max U_{Res}(\text{“}H_0 \text{ is T”}|H_0 \text{ is F}) &= 1 \implies \\ \mathbb{E}_\theta[U_{\text{Res}}(\text{“}H_0 \text{ is F”}|H_0 \text{ is F})] &> \\ \mathbb{E}_\theta[U_{\text{Res}}(\text{“}H_0 \text{ is T”}|H_0 \text{ is F})\end{aligned}$$

The payoff from accurately reporting the state of the world when the null hypothesis is in fact False is larger than that of any achievable payoff from misreporting. Hence, misreporting only occurs when the null hypothesis is in fact True.

Proof of Lemma 2

Lemma 2. Replicator's I-C Constraint. *There is a cutpoint type* $\tilde{\theta}(\beta,\chi,\kappa)$ *such that, in any separating equilibrium, all types* $\theta \leq \tilde{\theta}(\beta,\chi,\kappa)$ *Replicate, and all types* $\theta > \tilde{\theta}(\beta,\chi,\kappa)$ *do not Replicate.*

Proof. Suppose that the Researcher reports "H_0 is F". Then, the Replicator's payoff from choosing to Replicate, given their beliefs, can be written as:

$$\begin{aligned}&\mathbb{E}_\omega[U_{\text{Rep}}(R|\text{“}H_0 \text{ is F”},\mu)] \\ &= (1-\beta)[U_{\text{Rep}}(R, m(T)=F)] + \beta[U_{\text{Rep}}(R, m(F)=F)] \\ =\ & (1-\beta)[1-\theta-\kappa] + \beta[\chi+\theta-\kappa]\end{aligned}$$

The Replicator's payoff from choosing not to Replicate is θ, irrespective of the Researcher's choice.

Rearranging, we have that:

$$\begin{aligned}\mathbb{E}_\omega[U_{\text{Rep}}(R|\text{“}H_0 \text{ is F”},\mu)] \geq \mathbb{E}_\omega[U_{\text{Rep}}(\neg R|\text{“}H_0 \text{ is F”},\mu)] &\iff \\ \theta \leq \frac{\beta\chi-\kappa}{2(1-\beta)} + \frac{1}{2} \\ \theta \leq \tilde{\theta}(\beta,\chi,\kappa)\end{aligned}$$

This describes the realized degree of preference alignment required for the Replicator to prefer not to replicate the paper.

Proof of Lemma 3

Lemma 3. Researcher's I-C Constraint. *There is a threshold* $\tilde{\rho}(\nu,\beta,\chi,\kappa)$ *such that, if* $\rho \geq \tilde{\rho}(\nu,\beta,\chi,\kappa)$*, the Researcher prefers to misreport when the null hypothesis is true.*

Proof. First, consider the cases where $Pr(\theta \leq \tilde{\theta}) \in \{0,1\}$. If $Pr(\theta \leq \tilde{\theta}) = 0$, then the Replicator never Replicates, and the Researcher always misreports. If $Pr(\theta \leq \tilde{\theta}) = 1$, then the Replicator always Replicates, and the Researcher never misreports. (This can be seen by inspecting the Replicator's Incentive-Compatibility Constraint.) The interesting case is then when $Pr(\theta \leq \tilde{\theta}) \in (0,1)$, in which case we have:

$$\begin{aligned}\mathbb{E}_\theta\left[m(T) = \text{“}F\text{”}\right] &\geq \mathbb{E}_\theta\left[m(T) = \text{“}T\text{”}\right] \\ (1+\nu)Pr[\theta > \tilde{\theta}(\beta,\chi,\kappa)] - Pr[\theta \leq \tilde{\theta}(\beta,\chi,\kappa)] &\geq 1 \\ \frac{\nu}{2+\nu} &\geq Pr(\theta \leq \tilde{\theta}) \\ \frac{\nu}{2+\nu} &\geq \tilde{\theta}(\beta,\chi,\kappa)^\rho \\ \rho \ln \tilde{\theta}(\beta,\chi,\kappa) &\leq \ln\nu - \ln(2+\nu)\end{aligned}$$

Note that since $\tilde{\theta}(\beta,\chi,\kappa) \in (0,1)$, $\ln \tilde{\theta}(\beta,\chi,\kappa) \in (-\infty, 0)$, so we have

$$\rho \geq \frac{\ln\nu - \ln(2+\nu)}{\ln \tilde{\theta}(\beta,\chi,\kappa)}$$

So that:

$$\begin{aligned}\mathbb{E}_\theta\left[m(T) = \text{“}F\text{”}\right] \geq \mathbb{E}_\theta\left[m(T) = \text{“}T\text{”}\right] &\iff \\ \rho \geq \tilde{\rho}(\nu,\beta,\chi,\kappa)\end{aligned}$$

As desired.

Proof of Lemma 4

Lemma 4. Readership's Strategy.

$$\beta < \frac{1}{2} \implies \mathbb{E}[U_{Reader}(\neg F|\neg R)] > \mathbb{E}[U_{Reader}(F|\neg R)]$$

Proof. Suppose the Readership plays a pure strategy and chooses $\neg F$ when $\beta < \frac{1}{2}$. Then their payoff is simply $(1)(1-\beta) + (-1)\beta = 1-2\beta$. Now, suppose that the Readership makes an $\epsilon-$deviation to playing $\neg F$ with probability $1-\epsilon$ and F with probability ϵ. The Readership's payoff becomes $1-2\beta+4\epsilon\beta-2\epsilon$. But since $\beta < \frac{1}{2}$, we have that $2\epsilon > 4\epsilon\beta$, and hence that $1-2\beta+4\epsilon\beta - 2\epsilon < 1-2\beta$. Hence, there is no profitable $\epsilon-$deviation from playing $\neg F$, and hence $\mathbb{E}[U_{Reader}(\neg F|\neg R)] > \mathbb{E}[U_{Reader}(F|\neg R)]$.

Proof of Proposition 1

Proposition 1. Collusive and Adversarial Replication.

Suppose Assumption 1 holds. Then:

1. *If Replicator is sufficiently socially connected to the Researcher, that is, when* $\rho \geq \tilde{\rho}(\nu,\beta,\chi,\kappa)$*, there is a pooling PBE in which the Researcher misreports their results, and a cutpoint type of the Replicator* $\tilde{\theta}(\beta,\chi,\kappa)$ *such that all Replicators of type* $\theta > \tilde{\theta}(\beta,\chi,\kappa)$ *choose not to Replicate. If no Replication occurs, the Readership chooses not to reject the null hypothesis; otherwise they accurately match their action to the state of the world.*
2. *Otherwise, there is a separating PBE in which the Researcher reports their results accurately, and all types of the Replicator replicate.*

Proof. This collects the results of Lemmas 1-4.

Proof of Lemma 5

Lemma 5.
$$FDR(\beta,\chi,\kappa,\rho) = (1-\beta)\left[1-\tilde{\theta}(\beta,\chi,\kappa)^{\rho}\right]$$

Proof.
$$\begin{aligned} FDR &= Pr(H_0 \text{ is True} \mid \text{“}H_0 \text{ is False”}, \neg R) \\ &= Pr(H_0 \text{ is True} \mid \text{“}H_0 \text{ is False”}) Pr(\neg R) \\ &= (1-\beta^{\text{post}}) Pr(\neg R) \\ &= (1-\beta) Pr[\theta > \tilde{\theta}(\beta,\chi,\kappa)] \\ &= (1-\beta)\left\{1 - Pr[\theta \le \tilde{\theta}(\beta,\chi,\kappa)]\right\} \\ &= (1-\beta)\left[1-\tilde{\theta}(\beta,\chi,\kappa)^{\rho}\right] \end{aligned}$$

Proof of Proposition 2

Proposition 2. Social Welfare Loss Due to Collusion.

i. There is a minimum readership size N_0 such that $N > N_0$ implies there is net social welfare loss from moving to the collusive equilibrium from the adversarial equilibrium.
ii. When $N > N_0$, the magnitude of the social welfare loss due to collusion is increasing in the False Discovery Rate.

Proof. To verify the first claim, we first find a *lower* bound for the social welfare loss, and show that it is *positive* when N is larger than some value N_0; i.e. that total welfare is lower in the collusive equilibrium for a large enough readership. Denoting:

$$\begin{aligned} Loss \equiv \sum_i & \mathbb{E}_{\theta,\beta}[U_i(s_i^*(\theta; s_{-i}^*(\theta)) | \rho < \tilde{\rho}(\nu,\beta,\chi,\kappa))] - \\ & \mathbb{E}_{\theta,\beta}[U_i(s_i^*(\theta; s_{-i}^*(\theta)) | \rho \ge \tilde{\rho}(\nu,\beta,\chi,\kappa))] \end{aligned}$$

Which we can write in terms of the individual utilities:

$$Loss = \sum_i \Delta U_i = \Delta U_{Res} + \Delta U_{Rep} + \Delta U_{Read}$$

Where:

$$\begin{aligned} \Delta U_i = & \mathbb{E}_{\theta,\beta}[U_i(s_i^*(\theta; s_{-i}^*(\theta)) | \rho < \tilde{\rho}(\nu,\beta,\chi,\kappa))] - \\ & \mathbb{E}_{\theta,\beta}[U_i(s_i^*(\theta; s_{-i}^*(\theta)) | \rho \ge \tilde{\rho}(\nu,\beta,\chi,\kappa))] \end{aligned}$$

Recall that $\rho < \tilde{\rho}$ characterizes the *adversarial equilibrium*, in which researchers do not misreport, and $\rho \ge \tilde{\rho}$ the *collusive equilibrium*, in which researchers do misreport. Hence each ΔU_i is the difference between the adversarial and the collusive equilibrium: when this is positive, welfare for player i is greater in the adversarial equilibrium.

We compute the losses for each player individually. It is sufficient to find lower bounds for the individual losses, since our goal is to bound social welfare loss from below: that is, we say that the social welfare loss is *at least* as great as some quantity. For the Researcher:

$$\begin{aligned} \Delta U_{Res} &= \mathbb{E}_{\theta,\omega}[U_{Res}(m^*(\omega)|\rho < \tilde{\rho})] - \\ &\quad \mathbb{E}_{\theta,\omega}[U_{Res}(m^*(\omega)|\rho \ge \tilde{\rho})] \\ &= (1-\beta)\left\{1 - Pr(\theta > \tilde{\theta})(1+\nu) - Pr(\theta \le \tilde{\theta})\right\} \\ &= (1-\beta) + (1-\beta)Pr(\theta \le \tilde{\theta}) - FDR(1+\nu) \\ &= (1-\beta) + [(1-\beta) - FDR] - FDR(1+\nu) \\ &= 2(1-\beta) - FDR(2+\nu) \\ &\ge -FDR(2+\nu) \end{aligned}$$

Notice that the Researcher's change in utility is negative, which reflects the fact that the Researcher is better off in the Collusive Equilibrium than the Adversarial Equilibrium. We find a lower bound for the the Replicator's loss from the bad equilibrium:

$$\begin{aligned} \Delta U_{Rep} &= \mathbb{E}_{\theta,\omega}[U_{Res}(R^*(\theta)|\rho < \tilde{\rho})] - \mathbb{E}_{\theta,\omega}[U_{Rep}(R^*(\theta)|\rho \ge \tilde{\rho})] \\ &= \int_0^1 (\chi + \theta - \kappa)\, d\theta \\ &\quad - \beta \int_0^1 \left(\mathbb{I}\{\theta \le \tilde{\theta}\}(\chi+\theta-\kappa) + \mathbb{I}\{\theta \ge \tilde{\theta}\}\theta\right) d\theta \\ &\quad - (1-\beta)\int_0^1 \left[\mathbb{I}\{\theta \le \tilde{\theta}\}(1-\theta-\kappa) + \mathbb{I}\{\theta \ge \tilde{\theta}\}\theta\right] d\theta \\ &= \int_0^1 (\chi+\theta-\kappa)\, d\theta - \beta\int_0^{\tilde{\theta}} (\chi+\theta-\kappa)\, d\theta \\ &\quad - (1-\beta)\int_0^{\tilde{\theta}} (1-\theta-\kappa)\, d\theta - \int_{\tilde{\theta}}^1 \theta\, d\theta \\ &= \beta\int_0^1 (\chi+\theta-\kappa)\, d\theta + (1-\beta)\int_0^1 (\chi+\theta-\kappa)\, d\theta \\ &\quad - \beta\int_0^{\tilde{\theta}} (\chi+\theta-\kappa)\, d\theta - (1-\beta)\int_0^{\tilde{\theta}} (1-\theta-\kappa)\, d\theta \\ &\quad - \int_{\tilde{\theta}}^1 \theta\, d\theta \\ &= \beta\int_0^1 (\chi+\theta-\kappa)\, d\theta + (1-\beta)\int_0^{\tilde{\theta}} (\chi+\theta-\kappa)\, d\theta \\ &\quad + (1-\beta)\int_{\tilde{\theta}}^1 (\chi+\theta-\kappa)\, d\theta - \beta\int_0^{\tilde{\theta}} (\chi+\theta-\kappa)\, d\theta \\ &\quad - (1-\beta)\int_0^{\tilde{\theta}} (1-\theta-\kappa)\, d\theta - \int_{\tilde{\theta}}^1 \theta\, d\theta \\ &= \beta\int_{\tilde{\theta}}^1 (\chi+\theta-\kappa)\, d\theta - (1-\beta)\int_0^{\tilde{\theta}} (1-\chi)\, d\theta \\ &\quad + (1-\beta)\int_{\tilde{\theta}}^1 (\chi+\theta-\kappa)\, d\theta - \int_{\tilde{\theta}}^1 \theta\, d\theta \\ &\ge -(1-\beta)\int_0^{\tilde{\theta}} (1-\chi)\, d\theta - \int_{\tilde{\theta}}^1 \theta\, d\theta \\ &\ge -(1-\beta)\frac{(1-\chi)}{2} - \frac{1}{2} \end{aligned}$$

So that:

$$\Delta U_{Rep} \ge -(1-\beta)\frac{(1-\chi)}{2} - \frac{1}{2}$$

The Readership's loss from collusion is:

$$\begin{aligned}\Delta U_{Read} &= N\left\{1 - [Pr(\theta \leq \tilde{\theta}) + Pr(\theta > \tilde{\theta})(1-2\beta)]\right\} \\ &= N\left\{1 - Pr(\theta \leq \tilde{\theta}) - Pr(\theta > \tilde{\theta})(1-2\beta)\right\} \\ &= N\left\{Pr(\theta > \tilde{\theta})2\beta\right\} \\ &= N\left\{FDR\frac{2\beta}{1-\beta}\right\} \\ &= FDR\left[N\frac{2\beta}{1-\beta}\right]\end{aligned}$$

A lower bound on welfare loss due to collusion is then:

$$\begin{aligned}\sum_i \Delta U_i &= \Delta U_{Res} + \Delta U_{Rep} + \Delta U_{Read} \\ &\geq FDR\left(N\frac{2\beta}{1-\beta} - 2 - \nu\right) - (1-\beta)\frac{(1-\chi)}{2} - \frac{1}{2} \\ &\equiv \sum_i \underline{\Delta} U_i\end{aligned}$$

So that $\sum_i \underline{\Delta} U_i \geq 0$ when:

$$\begin{aligned}FDR\left(N\frac{2\beta}{1-\beta} - 2 - \nu\right) - (1-\beta)\frac{(1-\chi)}{2} - \frac{1}{2} &\geq 0 \\ N\frac{2\beta}{1-\beta} - 2 - \nu &\geq FDR^{-1}\left[(1-\beta)\frac{(1-\chi)}{2} + \frac{1}{2}\right] \\ N &\geq \frac{1-\beta}{2\beta}\left\{FDR^{-1}\left[(1-\beta)\frac{(1-\chi)}{2} + \frac{1}{2}\right] + 2 + \nu\right\}\end{aligned}$$

Note that this depends only on exogenous parameters. Hence, define:

$$N_0 \equiv \left\lceil \frac{1-\beta}{2\beta}\left\{FDR^{-1}\left[(1-\beta)\frac{(1-\chi)}{2} + \frac{1}{2}\right] + 2 + \nu\right\}\right\rceil$$

Then:

$$\begin{aligned}N > N_0 &\implies 0 < \sum_i \underline{\Delta} U_i \leq \sum_i \Delta U_i \\ N > N_0 &\implies 0 < \sum_i \underline{\Delta} U_i \leq Loss\end{aligned}$$

In words, when the readership is larger than the threshold value N_0, welfare in the adversarial equilibrium exceeds that in the collusive equilibrium.

To show the second claim, first note that:

$$\frac{\partial \sum_i \underline{\Delta} U_i}{FDR} = N\frac{2\beta}{1-\beta} - 2 - \nu$$

And, second, that we can write N_0 as, for a positive constant α:

$$\left\lceil \frac{1-\beta}{2\beta}(\alpha + 2 + \nu)\right\rceil$$

So that $N > N_0 \implies$:

$$\frac{\partial \sum_i \underline{\Delta} U_i}{FDR} > \frac{2\beta}{1-\beta}\left\lceil \frac{1-\beta}{2\beta}(\alpha + 2 + \nu)\right\rceil - 2 - \nu > 0$$

Since this is the lower bound on the social welfare loss, this entails that increasing FDR leads to increasing social welfare loss, which is what we wanted to show.

Comparative Statics

Proposition 3.

i. The FDR is increasing in ρ (the degree of social connectedness), and the κ (the cost of replication).

ii. The FDR is decreasing in χ (the incentive to confirm existing studies) and β (the plausibility of the alternative hypothesis or the power of the statistical procedure).

iii. The threshold value of ρ needed to sustain the Collusive Equilibrium is decreasing in ν (the return to novelty).

Proof. The first two claims are straightforwardly verified by taking derivatives of the FDR with respect to the relevant parameters in the expression:

$$(1-\beta)\left[1 - \tilde{\theta}(\beta, \chi, \kappa)^\rho\right]$$

Where, as previously:

$$\tilde{\theta}(\beta, \chi, \kappa) = \frac{\beta\chi - \kappa}{2(1-\beta)} + \frac{1}{2}$$

So that the FDR can be written as:

$$(1-\beta)\left[1 - \left(\frac{\beta\chi - \kappa}{2(1-\beta)} + \frac{1}{2}\right)^\rho\right]$$

For the third claim, we have:

$$\tilde{\rho} = \frac{\ln \nu - \ln(2+\nu)}{\ln \tilde{\theta}}$$

Where:

$$\tilde{\rho} \propto \ln(2+\nu) - \ln(\nu)$$

Taking derivatives, we have:

$$\frac{\partial \tilde{\rho}}{\partial \nu} = \frac{1}{2+\nu} - \frac{1}{\nu}$$

So that:

$$\frac{\partial \tilde{\rho}}{\partial \nu} < 0$$

For any $\nu > 0$.

A.2: Random Auditing

Proposition 4.

Relative to the Misreporting game:

i. The threshold value $\tilde{\rho}(\gamma)$ is larger, meaning that a higher degree of social connectedness, and a minimum novelty premium $\tilde{\nu}(\gamma)$ required to sustain the collusive equilibrium.

ii. The FDR is smaller for any nonzero value of γ.

iii. The FDR is decreasing, and hence social welfare is increasing, in γ.

The Researcher's Incentive-Compatibility Constraint under Random Auditing becomes:

$$\mathbb{E}_{\theta,\gamma}\left[m(T) = \text{“}F\text{”}\right] \geq \mathbb{E}_{\theta,\gamma}\left[m(T) = \text{“}T\text{”}\right] \iff$$

$$
\begin{aligned}
(1+\nu)(1-\gamma)Pr(\theta > \tilde{\theta}) - [\gamma + (1-\gamma)Pr(\theta \leq \tilde{\theta})] &\geq 1 \\
\frac{(1+\nu)(1-\gamma) - (1+\gamma)}{(1+\nu)(1-\gamma) + (1-\gamma)} &\geq Pr(\theta \leq \tilde{\theta}) \\
\frac{(1+\nu)(1-\gamma) - (1+\gamma)}{(1+\nu)(1-\gamma) + (1-\gamma)} &\geq \tilde{\theta}^{\rho}
\end{aligned}
$$

First note, that since $\theta^{\rho} \geq 0$, for this inequality to be satisfied we need:

$$
\nu \geq 2\left(\frac{\gamma}{1-\gamma}\right) \equiv \tilde{\nu}(\gamma)
$$

Otherwise the Researcher never misreports, and there is no equilibrium in collusion. We next suppose that this is the case. We then have, taking logs:

$$
\begin{aligned}
\rho \ln \tilde{\theta} &\leq \ln[(1+\nu)(1-\gamma) - (1+\gamma)] - \ln[(1+\nu)(1-\gamma) + (1-\gamma)] \\
\rho &\geq \frac{\ln[(1+\nu)(1-\gamma) - (1+\gamma)] - \ln[(1+\nu)(1-\gamma) + (1-\gamma)]}{\ln \tilde{\theta}} \\
\rho &\geq \tilde{\rho}(\nu, \beta, \chi, \kappa, \gamma)
\end{aligned}
$$

Next we want to show that this is larger than the degree of social connectedness required in the Misreporting Game. We have that:

$$
\begin{aligned}
0 < \frac{\nu - 2\gamma - \nu\gamma}{(2+\nu) - 2\gamma - \nu\gamma} &< \frac{\nu}{2+\nu} \\
\ln[\nu - 2\gamma - \nu\gamma] - \ln[(2+\nu) - 2\gamma - \nu\gamma] &< \ln \nu - \ln(2+\nu) < 0 \\
\frac{\ln[\nu - 2\gamma - \nu\gamma] - \ln[(2+\nu) - 2\gamma - \nu\gamma]}{\ln \tilde{\theta}} &> \frac{\ln \nu - \ln(2+\nu)}{\ln \tilde{\theta}} > 0 \\
\tilde{\rho}(\nu, \beta, \chi, \kappa, \gamma) &> \tilde{\rho}(\nu, \beta, \chi, \kappa) > 0
\end{aligned}
$$

As desired; or, in words, that the level of social connectedness between the Researcher and the Replicator pool required to sustain an equilibrium in misreporting is larger under random auditing.

The FDR becomes:

$$
(1-\beta)(1-\gamma)[1 - Pr(\theta \leq \tilde{\theta})]
$$

Which is smaller than the FDR in the previous setting, for any $\gamma \in (0, 1]$, and is decreasing in γ. As $\gamma \to 1$, we have $FDR \to 0$, and as $\gamma \to 0$, we have $FDR(\gamma, \beta, \chi, \kappa, \rho) \nearrow FDR(\beta, \chi, \kappa, \rho)$. Hence we have that:

$$
\frac{\partial FDR(\gamma, \beta, \chi, \kappa, \rho)}{\partial \gamma} < 0
$$

Or that increasing the probability of a random audit decreases the FDR. Since we have already established that social welfare is a decreasing function of the FDR, this establishes that social welfare is increasing in the probability of a random audit.

A.3: Noisy Replications

Proposition 5.

There exists a threshold value λ^*, such that:

- *i. When $\lambda \leq \lambda^*$ entails that Replication is uninformative, the Readership follows its prior belief in all circumstances, there is no social welfare loss from collusion as the research process does not inform the Readership's action.*
- *ii. When $\lambda > \lambda^*$, Replication informs the Readership's action, and there is positive social welfare loss from collusion.*
- *iii. The equilibrium described in Proposition 1 characterizes the case where $\lambda = 1$.*

Suppose that Replication is not perfectly informative, but instead, when academic fraud occurs, a Replication can detect it with probability λ.

This modification to the setup affects the incentives of players. First, it affects Replicator incentives: more-aligned Replicators are now more likely to Replicate, because they are less likely to detect fraud. Second, it affects Researcher incentives: more types of the Replicator choose to Replicate, but are less likely to detect fraud.

Replicator's IC Constraint The Replicator now prefers to replicate if:

$$
\begin{aligned}
\theta &\leq (1-\beta)\left[\lambda(1-\theta-\kappa) + (1-\lambda)(\chi+\theta-\kappa)\right] \\
\theta &\leq \frac{(1-\beta)\left[\lambda + (1-\lambda)\chi\right] + \beta\chi - \kappa}{2\lambda(1-\beta)} \\
\theta &\leq \frac{1}{2} + \left[\frac{1-\lambda}{\lambda} + \frac{\beta}{(1-\beta)}\right]\frac{\chi}{2} - \frac{\kappa}{2\lambda(1-\beta)}
\end{aligned}
$$

This expression is proportional to $\lambda^{-1}(\chi - \kappa)$, where, by Assumption 1, $(\chi - \kappa) > 0$. Hence it is decreasing in λ, so that $\tilde{\theta}(\lambda, \beta, \chi, \kappa) > \tilde{\theta}(\beta, \chi, \kappa)$, for $\lambda \in (0, 1)$. Intuitively, when the probability of detecting fraud decreases, more types of the Replicator are willing to Replicate.

Note that when $\lambda = 1$, this expression is equal to:

$$
\frac{\beta\chi - \kappa}{2(1-\beta)} + \frac{1}{2}
$$

as above.

Researcher's IC Constraint When $\lambda \in (0, 1)$, the Researcher prefers to misreport when:

$$
\begin{aligned}
Pr(\theta \leq \tilde{\theta})(1+\nu) + Pr(\theta > \tilde{\theta})\left[(1-\lambda)(1+\nu) - \lambda\right] &\geq 1 \\
Pr(\theta \leq \tilde{\theta}) &\leq \frac{\nu}{\lambda(2+\nu)}
\end{aligned}
$$

Note that when $\lambda = 1$, this is the same result in the previous model. The expression on the right-hand side is decreasing in λ, so that, in general, the Researcher is more willing to misreport when Replication is less likely to reveal information about the state of the world.

The minimum value of $\tilde{\rho}(\lambda)$ needed to sustain the Collusive Equilibrium then becomes:

$$
\tilde{\rho}(\lambda) \geq \frac{\ln \nu - \ln \lambda(2+\nu)}{\ln \tilde{\theta}(\lambda, \beta, \chi, \kappa)}
$$

Note that the numerator decreases slower than linearly in λ, while the denominator decreases faster than linearly in λ, so that $\tilde{\rho}(\lambda)$ is increasing in λ. This means that $\tilde{\rho}(\lambda) \nearrow \tilde{\rho}$ as $\lambda \nearrow 1$.

Substantively, as Replication becomes less informative, a smaller degree of social connectedness is needed to sustain the equilibrium in misreporting. This is intuitive, as it entails that the less effective auditing is, the less incentive-alignment is needed between the Researcher and the Replicator to sustain an equilibrium in collusion. It is easier to 'get away' with bad research practices.

Reader's Beliefs Next, by Bayes' rule, we can state the probability that the null hypothesis is False given a Replication that reports that the null is False:

$$\begin{aligned} Pr(H_0 \text{ is F}|R = F) &= \frac{Pr(R = F|H_0 \text{ is F})Pr(H_0 \text{ is F})}{Pr(R = F)} \\ &= \frac{\lambda\beta}{\lambda\beta + (1-\lambda)(1-\beta)} \\ &= \beta^* \end{aligned}$$

Note that $\lambda > \frac{1}{2} \implies \beta^* > \beta$, so that updating in the correct direction occurs. It then suffices to study the case where $\lambda \in (\frac{1}{2}, 1)$.

By Lemma 4, we know that $\beta = \frac{1}{2}$ is the posterior belief threshold that induces the Readership to endorse the conclusion that the null is false (to take the action congruent with the state of the world).

There is a threshold level of precision of Replication, λ^*, such that $\lambda > \lambda^*$ implies that the Readership would prefer to follow the advice of the Replicator, rather than follow their prior beliefs. In particular, this is the degree of belief updating required, given some actual prior value β, to induce the Readership to take a different action given the Replicator's report.

We can solve for this value as follows. Let $\beta = \frac{1}{2} - c$, for some constant $c \in \left(0, \frac{1}{2}\right)$. (Recall that we supposed $\beta < \frac{1}{2}$ in Assumption 1, to ensure that the default action taken by the Readership was not to believe the Researcher's report.)

Next note that, by Lemma 4, the Researcher endorses the conclusion that the null is False when $Pr(H_0 \text{ is F} \,|R = F) > \frac{1}{2}$. We can write the Readership's posterior as follows, and solve for λ^* in terms of c:

$$Pr(H_0 \text{ is F} \,|R = F) = \frac{\lambda(\frac{1}{2} - c)}{\lambda(\frac{1}{2} - c) + (1-\lambda)(\frac{1}{2} + c)}$$

$$\begin{aligned} Pr(H_0 \text{ is F} \,|R = F) > \frac{1}{2} &\iff \\ \frac{1}{2} &< \frac{\lambda^*(\frac{1}{2} - c)}{\lambda^*\left(\frac{1}{2} - c\right) + (1-\lambda^*)\left(\frac{1}{2} + c\right)} \\ 2\lambda^*\left(\frac{1}{2} - c\right) &> \lambda^*\left(\frac{1}{2} - c\right) + (1-\lambda^*)\left(\frac{1}{2} + c\right) \\ \lambda^*\left(\frac{1}{2} - c\right) &> (1-\lambda^*)\left(\frac{1}{2} + c\right) \\ \lambda^* &> \frac{1+4c}{2+2c} \end{aligned}$$

Note that when $c = 0$, $\beta = \frac{1}{2}$, and any $\lambda > \frac{1}{2}$ is sufficient to induce the Readership to take the correct action.

Note that for $\lambda < \lambda^*$, the Readership would prefer to follow their prior beliefs about the state.

Intuitively, this entails that the Readership prefers to follow the advice of the Replicator only when Replication is sufficiently accurate.

Social Welfare We now have two cases to study.

First, when $\lambda < \lambda^*$, the Readership simply follows its prior, no matter what the Replicator does, in which case they earn $N\{(1)(1-\beta) - 1(\beta)\} = N(1-2\beta)$. Readership welfare does not differ across equilibria, in this instance, since Replication does not inform the Readership's action.

When $\lambda > \lambda^*$, the Readership's welfare loss from collusion becomes:

$$\begin{aligned} \Delta U_{Read}|\lambda > \lambda^* &= N\Big\{(2\lambda - 1) - Pr(\theta \le \tilde{\theta})[2\lambda - 1] \\ &\quad - Pr(\theta > \tilde{\theta})[1 - 2\beta]\Big\} \\ &= N\Big\{[1 - Pr(\theta \le \tilde{\theta})][2\lambda - 1] \\ &\quad - Pr(\theta > \tilde{\theta})[1 - 2\beta]\Big\} \\ &= N\left\{[1 - Pr(\theta \le \tilde{\theta})][2\lambda + 2\beta - 2]\right\} \end{aligned}$$

As shown above, $Pr(\theta \le \tilde{\theta})$ is decreasing in λ, which entails that Readership welfare from moving to the adversarial equilibrium is increasing in λ, given that $\lambda > \lambda^*$.

That is, we can rank the equilibria by the degree of information revelation that occurs in each: Readership welfare (and hence social welfare) in any Collusive Equilibrium is increasing in the informativeness of Replication. The Adversarial Equilibrium, which induces maximum information revelation for any given level of λ, maximizes social welfare.

To summarize: when Replication is not perfectly informative, there is a threshold value of λ, λ^*, such that $\lambda < \lambda^*$ entails that no learning about the state occurs in equilibrium, the Research process is uninformative, and there is no harm to social welfare caused from collusion between Researcher and Replicator (since Replication is insufficiently informative to aid the Replicator in making a decision about the state of the world). Otherwise, when Replication is sufficiently informative, there is social loss from the Collusive Equilibrium that is increasing in the informativeness of Replication. As Replication becomes more informative, the Readership loses out to a greater extent by Replicators' failures to hold Researchers accountable.