Collusive and Adversarial Replication

Research and Politics
XX(X):1–12

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

Anonymous Author[s]

Abstract

We describe a game in which social ties between Researchers and Replicators may prevent prospective replicators from debunking papers that misreport the state of the world. In this game, replication is an entrance decision, as a Replicator chooses whether or not to Replicate a given paper. 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 ultimately improves social welfare. We study three proposals to improve replication practices: Random auditing, or Police-Patrol Replication; automated unit tests, and lowering 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

The issue of replicability has drawn increasing attention, as 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 phishing, and site selection may or may not be caught be peer reviewers. Replicability is also tied to the generalizability of a phenomenon, and cumulative learning about a process of fundamental interest. 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 understudied topic.

We study Replication in an *entrance* model, in which Replicators can choose to opt in to performing Replications. We highlight the role of preference alignment between Researchers and prospective Replicators as a key determinant of whether or not Replication actually occurs.

We highlight *the role of social connectivity in the publication process*, which is a source of bias that is infrequently discussed in the literature. 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 discipline-wide False Discovery Rate (Benjamini and Hochberg 1995; Storey 2003), as examples of misreporting go unchecked. This harms the reading public, and hence the FDR is a measure of social welfare.

We then model the impact of random auditing, or Police-Patrol Replication, on researcher incentives, and show that centralized replication practices can reduce the FDR, our measure of social welfare in a hypothesis testing setting. We also discuss the adoption of automated unit testing, and the possible impact of a recent proposal to reduce the significance threshold from $p \leq .05$ to $p \leq .005$, showing that its effects on researcher incentives are ambiguous. 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. While our model specifically describes misreporting under fraud, we believe its morals generalize to other examples of poor academic behaviour, and can help to explain situations in which a discipline tolerates suboptimal research practices.

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 body of contemporary work in selective inference, in which Researcher decisions can be modelled as statistical features of a hypothesis testing problem (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 (Rosenthal 1979; Fithian et al. 2017; Bates et al. 2022, 2023; Wang et al. 2023; Andrews and Shapiro 2021; Frankel and Kasy 2022).

Within Political Science, Gerber et al. (2001), Gerber and Malhotra (2008), and Owen and Li (2021) provide evidence of the existence of publication bias within political science. Despite consensus on the importance of replication to a healthy publication process, the incentives of replicators are an understudied topic. Berinsky et al. (2021) model several sources of selection bias: file drawer bias, a preference for novelty, and a 'gotcha' or 'overturn' bias, in which replicator prefers to replicate a study when the result is likely to be

Anonymous Affiliation

Corresponding author:

easy to falsify. Galiani et al. (2017) study the problem of reproducibility in economics, and find empirical evidence that journals are less likely to publish replications that do not overturn existing findings, or for less 'important' papers.

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 benefit from accurate papers and are harmed by inaccurate papers.

Order of actions

1. Nature chooses the state of the world:
$$\omega = H\{H_0 \text{ is True}\} + (1 - H)\{H_0 \text{ is False}\}$$
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 Replicate (R) the paper, and hence report the correct state of the world, or not ($\neg R$).
4. 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, 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. This is analogous to an altruism parameter: social connectedness between academics implies a degree of 'linked fate', or investedness in each another's careers. The Replicator earns $\chi > 0$ if they *confirm* an existing finding; that is, if they choose to Replicate, and the Researcher did not misreport the state of the world. The Replicator pays a fixed *cost* $\kappa > 0$ if they choose to Replicate. Finally, the Replicator earns 1 if they debunk the paper. Hence a researcher faces a conflict between their connectedness to the Researcher, and the professional rewards to debunking a paper.

The Readership's Payoffs The Readership is composed of N readers who each earn 1 if either the Researcher's report matches the state of the world, or the Researcher's false report is debunked (and hence they are able to update their beliefs correctly given the Replicator's report). Readers each earn -1 otherwise.

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. In what follows, we characterize the conditions that sustain an equilibrium in misreporting.

Solutions

To simplify notation, we write the states of the world to $\omega = \{T, F\}$, and we describe a message function $m : \Omega \rightarrow$ "Ω".§

We restrict our attention to parameter values in which 'interesting' equilibrium behaviour occurs (we describe equilibria in the absence of this assumption in the Appendix).

Assumption 1. *$\nu > 0$ and $1 > \chi > \kappa > 0$.*

This ensures that the novelty premium is non-zero, and that there is an incentive to perform confirmatory replications of studies.

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.

*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).

†Though we do not model this explicitly, we can think of this as the connectivity of a graph that describes the discipline's social and professional network (Diestel 2017).

‡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.

§We will sometimes write, e.g. $m(T) =$ "F" to describe the event that the researcher reports that "H_0 is False" when H_0 is in fact true.

Replicator's Beliefs Let β^{post} be the Replicator's posterior belief that the null hypothesis is false, given the Researcher's report. By Bayes' rule, have that:

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

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 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 sharing in 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:

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

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.

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 Reviewer $\tilde{\theta}(\beta, \chi, \kappa)$ such that all Replicators of type $\theta > \tilde{\theta}(\beta, \chi, \kappa)$ choose not to Replicate.*
2. ***Adversarial Equilibrium.** Otherwise, there is a separating PBE in which the Researcher reports their results accurately, and all types of the Replicator replicate.*

When the pool of Replicators is sufficiently homogeneous, 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 than 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 is a statistic that summarizes 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; Bates et al. 2023). 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*

$$FDR \equiv Pr(H_0 \text{ is True} \,|\, \text{“}H_0 \text{ is False”}, \neg R)$$

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 4.

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

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.

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 positive social welfare loss in the collusive 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.

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).*

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 Reviewer 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. Replication should be tied to career incentives

Costs of Replication

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

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

¶This can be understood as an endogenous gotcha premium (Berinsky et al. 2021).

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 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 the Appendix.

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 moral is that random auditing can crowd out the possibility of misreporting induced by the selective entrance of Replicators. A higher level of discipline-wide social connectivity is required to sustain the collusive equilibrium, in order to counterbalance the effect of random auditing.

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 characterizes 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). A high degree of social alignment between Researchers and Replicators can therefore help to sustain cultural norms that are permissive of suboptimal research practices.

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.

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 Acceptance Tests

In technology companies engaged in software engineering, a standard part of the workflow to use unit 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 an 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 Significance Threshold

The results in Proposition 3 bear relevantly on a recent proposal to reduce the threshold for significance from .05 to .005 (Benjamin et al. 2021). 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. Since this is not the primary focus of this article, we do not assert that this

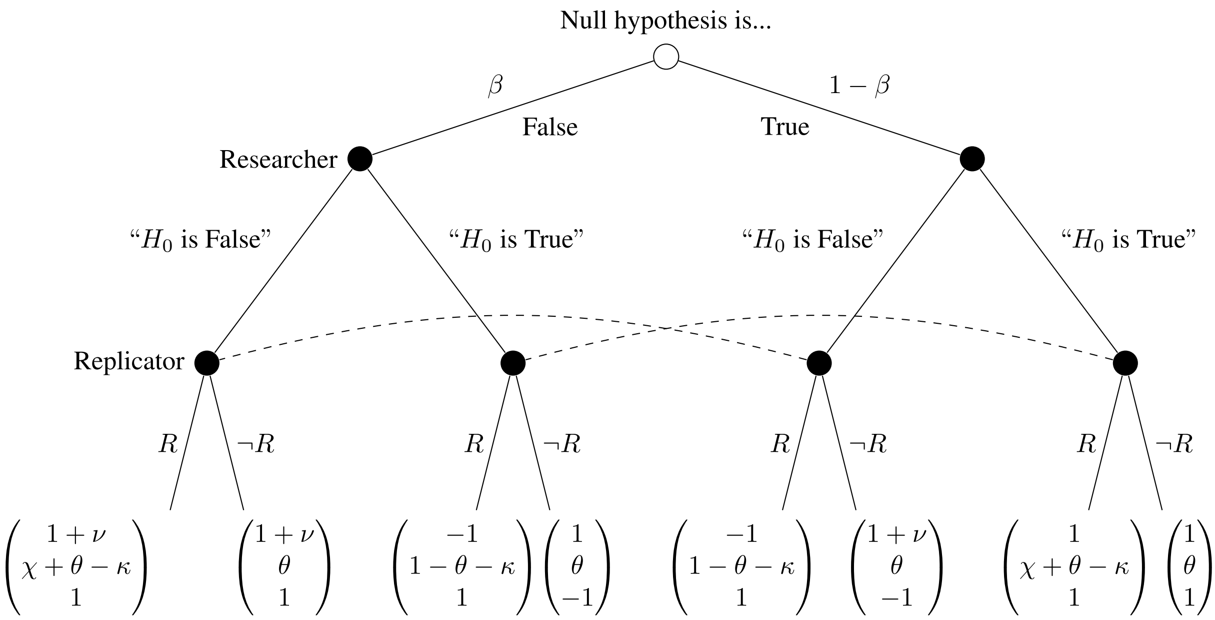

Figure 1. The Misreporting Game

would necessarily occur. Instead, the moral is that it is not sufficient to consider the statistical implications of adjusting the conventional threshold for significance alone: its effect on researcher incentives must 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.

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.

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 EJ, Al-Khatib SM, Desvigne-Nickens P, Alonso A, Djoussé L, Forman DE, Gillis AM, Hendriks JML, Hills MT, Kirchhof P, Link MS, Marcus GM, Mehra R, Murray KT, Parkash R, Piña IL, Redline S, Rienstra M, Sanders P, Somers VK, Wagoner DRV, Wang PJ, Cooper LS and Go AS (2021) Research priorities in the secondary prevention of atrial fibrillation: A national heart, lung, and blood institute virtual workshop report. *Journal of the American Heart Association* 10(16). DOI:10.1161/jaha.121.021566. URL https://doi.org/10.1161/jaha.121.021566.

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.

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.

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.

Galiani S, Gertler P and Romero M (2017) Incentives for replication in economics. Working Paper 23576, National Bureau of Economic Research. DOI:10.3386/w23576. URL http://www.nber.org/papers/w23576.

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.

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.

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.

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.

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

Proofs of Results

Proof of Lemma 1

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

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

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:

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

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

Rearranging, we have that:

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

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

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:

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

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

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

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.

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

Proof.

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

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 Reviewer $\tilde{\theta}(\beta, \chi, \kappa)$ such that all Replicators of type $\theta > \tilde{\theta}(\beta, \chi, \kappa)$ choose not to Replicate.*
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 1-3.

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 positive social welfare loss in the collusive 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 $\tilde{N}$ is larger than some value N_0. Denoting:

$$Loss \equiv \sum_i \mathbb{E}_{\theta,\beta}[U_i(s^*(\theta)|\rho < \tilde{\rho}(\nu,\beta,\chi,\kappa))] - \mathbb{E}_{\theta,\beta}[U_i(s^*(\theta)|\rho \geq \tilde{\rho}(\nu,\beta,\chi,\kappa))]$$

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:

$$\Delta U_i = \mathbb{E}_{\theta,\beta}[U_i(s^*(\theta)|\rho < \tilde{\rho})] - \mathbb{E}_{\theta,\beta}[U_i(s^*(\theta)|\rho \geq \tilde{\rho})]$$

Recall that $\rho < \tilde{\rho}$ characterizes the *adversarial equilibrium*, in which researchers are not socially connected, and $\rho \geq \tilde{\rho}$ the *collusive equilibrium*, in which researchers are socially connected.

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.[‖] We have:

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

We find a lower bound for the the Replicator's loss from the bad equilibrium:

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

So that:

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

The Readership's change in welfare is:

$$\begin{aligned}
\Delta U_{Read} &= \tilde{N} - [(1 - FDR)\tilde{N} - FDR\tilde{N}] \\
&= \tilde{N} - (1 - 2FDR)\tilde{N}2(1-\beta) - FDR(2+\nu) \\
&= 2FDR\tilde{N}
\end{aligned}$$

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

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

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

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

[‖]Noting that $FDR = (1-\beta) - (1-\beta)\tilde{\theta}^{\rho}$ implies $(1-\beta)\tilde{\theta}^{\rho} = (1-\beta) - FDR$

Where, since $FDR \in [0, 1-\beta]$, we have that this is positive when:

$$\tilde{N} \geq \frac{(1-\beta)\left[(1-\beta)(1-\chi)+\frac{1}{2}\right]}{2} + 1 + \frac{\nu}{2}$$

Hence, define:

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

Then:

$$\tilde{N} > N_0 \implies 0 < \sum_i \underline{\Delta} U_i \leq \sum_i \Delta U_i$$
$$\tilde{N} > N_0 \implies 0 < \sum_i \underline{\Delta} U_i \leq Loss$$

As desired.

To show the second claim, first note that:

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

And that:

$$N > N_0 \implies 2N > 2 + \nu$$

So that:

$$N > N_0 \implies \frac{\partial \sum_i \underline{\Delta} U_i}{FDR} > 0$$

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:

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

So that the FDR can be written as:

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

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 $\nu > 0$.

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:

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

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

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

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

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:

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

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

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

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.