# Examining the Role of Incentives in Scholarly Publishing with Multi-Agent Reinforcement Learning

Giulia Bernardi[1], Eric Medvet[2], Alberto Bartoli[2], and Andrea De Lorenzo[2]

[1]Department of Mathematics and Geosciences, University of Trieste, Italy
[2]Department of Engineering and Architecture, University of Trieste, Italy
giulia.bernardi@studenti.units.it, emedvet@units.it, bartoli.alberto@units.it, andrea.delorenzo@units.it

Downloaded from http://mitp.silverchair.com/isal/proceedings-pdf/isal2023/35/121/2355001/isal_a_00585.pdf by guest on 16 September 2024

## Abstract

Scientific research plays a crucial role in advancing human civilization, thanks to the efforts of a multitude of individual actors. Their behavior is largely driven by individual incentives, both explicit and implicit. In this paper, we propose and validate a multi-agent model to study the complex system of scholarly publishing and investigate the impact of incentives on research output. We use reinforcement learning to make the behavior of the actors optimizable, and guide their optimization with a reward signal that encodes the incentives. We consider various combinations of incentives and predefined behaviors and analyze their impact on both individual (h-index, impact factor) and overall indexes of research output. Our results suggest that, despite its simplicity, our model is able to capture the main dynamics of the system. Moreover, we find that (a) most incentives tend to favor productivity over quality and (b) incentives related to journal perceived reputation tend to result in waste of research efforts.

## Introduction

Scholarly publishing is an essential component of scientific research and involves interactions among different actors—authors, reviewers, editors, publishing companies—whose individual behavior is shaped by different types of incentives, explicit or implicit. It is clearly of the uttermost importance that the behavior emerging from such a system guarantees the scientific quality of the published works. On the other hand, there are many opportunities for individual actors to behave in such a way that their respective incentives are maximized without resulting in overall scientific progress. Indeed, the research community as a whole is increasingly concerned with such forms of questionable and often plainly fraudulent behaviors (Bartoli et al., 2016; Byrne et al., 2019; Bartoli and Medvet, 2020; Abalkina, 2021; Cabanac et al., 2021).

In this work we pursue the ambitious objective of investigating how incentives of single actors may impact the output of the overall scholarly publishing system. To this end we propose and assess experimentally, by means of simulation, a *multi-agent* model where the behavior of each agent is progressively optimized through *reinforcement learning*, that is, by means of a reward signal in which we encode the respective incentive. We analyze several different scenarios and analyze the resulting output indexes both on the system as a whole (e.g., amount and aggregate quality of published works) and at the level of single actors (e.g., author h-index or journal impact factor).

Although this exploratory work is clearly not sufficient to draw any firm conclusions regarding the impact of individual incentives on the emergent behavior of such a complex system as scholarly publishing, our results appear to be able to capture the main dynamics of the system. In particular, they seem to confirm the widespread view that current incentives tend to favor productivity over quality (Bartoli and Medvet, 2014; Ruocco et al., 2017; Baccini et al., 2019). Moreover, the results suggest that some authors' incentives, namely the one based on the reputation of the journal, tend to result in a waste of scientific production.

## Related Work

Due to its complexity, the scholarly publishing system has many potential issues: Rahal et al. (2023) listed those related to funding and race to publication venue prestige and showed that they often lead to questionable research practices and promote small, catchy papers instead of carefully planned complex investigations. To examine the shortcomings of the scholarly publishing system, there are two possible approaches: to measure the quality using bibliometric indicators (Bornmann and Mutz, 2015) or to simulate the academic publication process.

Many author-level bibliometric indicators have been proposed, including the h-index, which captures productivity and citation impact (Hirsch, 2005) and has been found to be consistent with peer judgment (Van Raan, 2006). However, criticisms have been made toward the practice of assessing authors mainly or solely by means of bibliometric indicators (Ruocco et al., 2017), including the fact that indicators may be used in a way that does not reflect their actual meaning (Hicks et al., 2015; Corrall et al., 2013; Barnes, 2017), or may lead to indicator-based behaviors—e.g., Baccini et al. (2019); Peroni et al. (2020) revealed how researchers gamed indicators through self-citations.

Two works investigated critical aspects of the academic publication process through simulation. Smaldino and McElreath (2016) presented an evolutionary model that show how the overemphasis on writing as many articles as possible can lead to the publication of false scientific discoveries. Medo and Cimini (2016) evaluated the validity of bibliometric indexes as a representation of authors. These articles provide insight into the potential dangers of certain academic incentives and the effectiveness of certain evaluation methods in the scholarly publishing system.

## Model

We model the scholarly publishing system as a multi-agent discrete-time system with two kinds of agents, authors and editors, and two entities, papers and journals. Authors produce papers over time and may submit them to journals. Editors are statically associated with journals, in a one-to-one relation, and evaluate papers as soon as they are submitted to their journal—for clarity, we will use the term journal for denoting both the entity and the associated agent, i.e., the editor.

Formally, we denote by $A$ the set of authors, by $P$ the set of papers, and by $J$ the set of journals. When the system evolves over time $A$ and $J$ are immutable, i.e., they never change, whereas $P$ may change.

### Authors, journals, and papers

An *author* $a \in A$ is an agent defined by two immutable attributes, the author's quality $q(a) \in [0, q_{\max}]$ and the author's productivity $\delta q(a) \in [\delta q_{\min}, \delta q_{\max}]$, and by two mutable attributes, the author's working paper $p_{\text{draft}}(a)$ and the set $P_A(a) \subseteq P$ of papers published by the author.

A *journal* $j \in J$ is an agent defined by one mutable attribute, the set $P_J(j) \subseteq P$ of papers published in that journal.

A *paper* $p$ is an entity with the following mutable attributes: the current quality $q(p) \in [0, q_{\max}]$, the maximum quality $q_{\max}(p) \in [0, q_{\max}]$, the writing time $t(p) \in \mathbb{N}$, the number $n_{\text{rej}}(p) \in \mathbb{N}$ of rejections, the references $R(p) \subseteq P$, and the publication year $y(p) \in \mathbb{N} \cup \varnothing$, where $\varnothing$ means that the paper has not yet been published.

### System evolution over time

Our system evolves over time as a consequence of the actions of authors and journals. However, the two kinds of agent perform their actions according to two different timings. Each author performs exactly one action at each time step. Each journal performs zero or more actions at each time step, once for each paper submitted to the corresponding journal.

We here describe how the system changes upon authors' and journals' actions. We will describe later how the agents choose which action to perform, i.e., what are the agents' policies.

**Author's actions.** At each $k$, an author $a$ performs one of the following $2 + |J|$ actions: *Restart* writing, *Keep* writing, or submit to a journal $j$ (briefly, *Submit-to-$j$*).

If $a$ performs the Restart action, we set the author's working paper $p' = p_{\text{draft}}(a)$ to a new paper. In detail, we set $q_{\max}(p') := [\mathcal{N}(q(a), \sigma_q)]_{0, q_{\max}}$, i.e., we sample a normal distribution with parameters $q(a)$, $\sigma_q$ (the latter being a parameter of the model), clamp the value to the proper interval $[0, q_{\max}]$, and assign it to $q_{\max}(p')$. We set $q(p') := [\mathcal{N}(\delta q(a), \sigma_{\delta q})]_{0, q_{\max}(p')}$, with $\sigma_{\delta q}$ being a parameter of the model. We set the other attributes, with the exception of the references, to 0 or $\varnothing$, i.e., $t(p) := 0$, $n_{\text{rej}}(p) := 0$, and $y(p) := 0$. For the references $R(p)$, we set them by sampling $P$, extracting papers with a probability proportional to their recency and quality until we collected $n_{\text{ref}}$ references, $n_{\text{ref}}$ being a parameter of the model. Namely, the probability of each paper $p$ to become a reference of $p'$ is proportional to the sum of normalized quality and recency:

$$\Pr(p) \propto \left( \underbrace{\frac{q(p)}{q_{\max}}}_{\text{quality}} + \underbrace{\left[ 1 - \frac{\lceil \frac{k}{12} \rceil - y(p)}{y_{\max}} \right]_{0,1}}_{\text{recency}} \right)^{\gamma_{\text{ref}}},$$

where $\lceil \frac{k}{12} \rceil$ is the current year and $y_{\max}$, $\gamma_{\text{ref}}$ are parameters of the models—the larger $\gamma_{\text{ref}}$, the stronger the preference for better and more recent papers. Intuitively, hence, upon the Restart action, the new working paper has an initial quality which depends on the author's productivity, a maximum quality that depends on the author's quality, and a fixed number of references selected depending on quality and recency. Note that we do not add the new working paper $p'$ to $P$ (hence it is not citable).

If the author $a$ performs the Keep action, we update the author's working paper $p' = p_{\text{draft}}(a)$. In detail, we set $q(p') := [q(p') + \mathcal{N}(\delta q(a), \sigma_{\delta q})]_{0, q_{\max}(p')}$ and $t(p') := t(p') + 1$. Intuitively, upon the Keep action, the working paper quality increases depending on the author's productivity.

Finally, if the author $a$ performs the action Submit-to-$j$, the corresponding journal agent $j$ is triggered and the outcome of $a$ action depends on $j$ action, as detailed in the next section.

**Journal actions.** Whenever an author $a$ with the working paper $p'$ performs the action Submit-to-$j$, $j$ performs one of the two following actions: *accept* or *reject*.

If $j$ performs the accept action, we set the paper publication year $y(p') := \lceil \frac{k}{12} \rceil$ and add $p'$ to the published papers sets, i.e., for the author $P_A(a) := P_A(a) \cup \{p'\}$, for the journal $P_J(j) := P_J(j) \cup \{p'\}$, and overall $P := P \cup \{p'\}$. Moreover, we set $a$ working paper to a new paper as for the case of the Restart author's action. Intuitively, hence, upon the action combination Submit-to-$j$ and accept, the paper

Downloaded from http://mitp.silverchair.com/isal/proceedings-pdf/isal2023/35/121/2355001/isal_a_00585.pdf by guest on 16 September 2024

is published and the authors starts working on a new paper. Note that, after publication, the paper attributes, never change, becoming in effect immutable.

If, otherwise, $j$ performs the reject action, we update the paper number of rejections $n_{\text{rej}}(p') := n_{\text{rej}}(p') + 1$. Intuitively, upon Submit-to-$j$ and reject, the paper remains a working paper.

### System initialization

We call *episode* an evolution of the system starting from time step $k = 0$ to time step $k = k_{\text{final}}$. At $k = 0$, we initialize the mutable and immutable attributes of all the authors, papers, and journals as follows.

For the authors, we compose $A$ by generating $n_{\text{authors}}$ authors, $n_{\text{authors}}$ being a parameter of the model. For each author $a$, we set the quality $q(a) := q_{\max}\mathcal{B}(\alpha_q, \beta_q)$ , i.e., we sample a Beta distribution with parameters $\alpha_q$, $\beta_q$ and multiply the value by $q_{\max}$, making it span in $[0, q_{\max}]$. In the experiments, we chose the values of $\alpha_q$, $\beta_q$ that correspond to a set of authors including many authors of "low quality" and few authors with "high quality". We set $a$ productivity $\delta q(a) := \delta q_{\min} + (\delta q_{\max} - \delta q_{\min})\mathcal{B}(\alpha_q, \beta_q)$, i.e., by sampling the Beta distribution and re-scaling the value to $[\delta q_{\min}, \delta q_{\max}]$. We set $a$ working paper to a new paper as for the case of the Restart author's action. We set $P_A(a) := \emptyset$.

For the journals, we simply compose $J$ by generating $n_{\text{journals}}$ empty journals, $n_{\text{journals}}$ being a parameter of the system. For each journal $j$, we set $P_J(j) := \emptyset$.

Finally, for the papers, we generate $n_{\text{papers}}$, $n_{\text{papers}}$ being a parameter of the model, as follows. For each $p$, we first select from $A$ an author $a$, to be the author of $p$, by sampling $A$ with a probability proportional to authors' productivity, and we update $P_A(a) := P_A(a) \cup \{p\}$. Then, we set the max quality $q_{\max}(p) := [\mathcal{N}(q(a), \sigma_{\delta q})]_{0,q_{\max}}$ and the quality $q(p) := q_{\max}(p)$. We set the year $y(p)$ to one of $0, -1, -2$, chosen randomly with equal probability. We set the $t(p) := 0$, $n_{\text{rej}}(p) := 0$, and $R(p) := \emptyset$. Once we compose $P$ with $n_{\text{papers}}$ papers as described above, we populate papers references and "publish" papers to journals as follows. For the references, we do as described above for the author's Restart action, for each paper—note that executing that procedure incrementally when creating paper, instead of all-at-once at the end, would have made the recency-quality random sampling less meaningful for the first papers, when $P$ is empty or contains few papers. For publishing papers to journals, we first sort the papers according to their quality, then partition them in $|J| = n_{\text{journals}}$ bins, and finally publish all the papers in a bin to the same journal $j$, i.e., for each $p$ in a bin, we update $P_J(p) := P_J(p) \cup \{p\}$.

### Limitations

We acknowledge that our model has several limitations. Firstly, there is no distinction among disciplines; all the authors and journals behave the same way independently of their field of research and all the papers can be referenced by all the other papers.

The authors, in our model, can only work on one project at a time and there is a simplistic relation between authors' effort and the resulting paper quality. Another limitation is that our authors' quality and productivity do not change overtime, leaving no space for career improvement or regression. Moreover, there is no co-authorship: each author works alone and is the sole driving force for advancing the corresponding working paper.

For what concerns journals, the editors' role in the scholarly publishing system is not taken into consideration, i.e., editors are disjoint from authors. Editors evaluate papers immediately, accepting or rejecting them; no effort is hence required for evaluating a paper and, in particular, the level of scrutiny is not related to the reviewing time. Additionally, for both journals and authors, institutions are not represented.

Overcoming these limitation would have significantly increased the complexity of our model. We selected the traits of the scholarly publishing system that we believe are more relevant for our study.

## Validation of the model

We performed an experimental evaluation for validating our model, i.e., for verifying that it is indeed capable of modeling the real scholarly publishing system.

To this end, we first defined some indexes useful in characterizing the scholarly publishing system. Then, we defined some policies governing the behavior of the modeled agents, i.e., authors and journals: we will call those policies *static* to distinguish them from the *learnable* ones that we will consider later. We created the static policies as a set of fixed rules that an author or a journal may realistically use, according to our experience. We set all the model parameters to represent at best the reality. And, finally, we performed many simulations, measured the indexes, and analyzed their behavior.

### Indexes

We consider two sets of indexes: a set of *global indexes* characterizing the entire scholarly publishing system and a set of *local indexes* characterizing individual agents in the system.

As global indexes, we define the following: the *total quality* $q_{\text{tot}} = \sum_{p \in P} q(p)$, where $P$ is the set of all published papers at the end of the episode; the *number* $|P|$ *of published papers*; the *rates* $\rho_{q,low}, \rho_{q,mid}, \rho_{q,hi}$ *of papers of low, medium, and high quality*, i.e., whose quality is in $\left[0, \frac{1}{3}q_{\max}\right[$, $\left[\frac{1}{3}q_{\max}, \frac{2}{3}q_{\max}\right[$, and $\left[\frac{2}{3}q_{\max}, q_{\max}\right[$, respectively; and the scholarly publishing system *efficiency* $e$. We define the efficiency as the ratio $e = \frac{q_{\text{tot}}}{q_{\text{tot,max}}}$ between the total quality $q_{\text{tot}}$ and the quality $q_{\text{tot,max}}$ that the system might have produced, i.e., if every authors' effort actually resulted in

Downloaded from http://mitp.silverchair.com/isal/proceedings-pdf/isal2023/35/121/2355001/isal_a_00585.pdf by guest on 16 September 2024

published papers: for the latter quantity, we consider the average productivity $\overline{\delta q} = \delta q_{\min} + (\delta q_{\max} - \delta q_{\min})\frac{\alpha_q}{\alpha_q+\beta_q}$ of the authors and multiply it by the number $|A|$ of authors and the duration $k_{\text{final}}$ of the episode, i.e., $q_{\text{tot,max}} = \overline{\delta q}|A|k_{\text{final}}$. We remark that in our model there are three sources of inefficiency, i.e., three kinds of waste of academic production: the first kind of waste occurs when authors work too long on a paper which has already reached the maximum quality (note that authors' quality and productivity are not related, hence low quality authors waste, in general, more); the second occurs when the author restarts a new paper, and hence drops the current one and every effort devoted to it; the third and last occurs when the author submits a paper which is (immediately) rejected. In the latter case, the author's productivity of the submission time step does not increase the working paper quality, nor is used in a new paper.

The above global indexes are meaningful and relevant to our study objectives, since they characterize the entire publishing system. However, a real-world counterpart of those indexes does not exist (maybe with the exception of $|P|$). It follows that we cannot use those indexes for quantitatively validating the model with real data. For overcoming this limitation, we also take in consideration two bibliometric local indexes which are widely used for assessing research.

Concerning authors, we consider the h-index (Hirsch, 2005), which is defined as the number $h$ of papers by the author having at least $h$ citations: we denote by $h(a)$ the h-index of an author $a$ at the end of an episode. We consider only the published papers when calculating $h(a)$.

For journals, we use the impact factor (Garfield, 1972). The impact factor $\text{IF}(j, y)$ of a journal $j$ at year $y$ is given by the number of citations received in year $y$ by the papers published in $j$ in the two preceding years divided by the number of such papers:

$$\text{IF}(j, y) = \frac{\sum_{p' \in P_y} |R_P(p') \cap (P_{J,y-1}(j) \cup P_{J,y-2}(j))|}{|P_{J,y-1}(j) \cup P_{J,y-2}(j)|},$$

where $P_y = \{p \in P : y(p) = y\}$ is the set of all papers published in year $y$ and $P_{J,y}(j) = P_J(j) \cap P_y$ is the set of papers published in year $y$ by journal $j$. We denote by $\text{IF}(j)$ the impact factor of a journal $j$ at the end of an episode.

## Static policies

We designed the static policies to mimic reasonable behaviors for authors or journals. Additionally, we designed some policies corresponding to random behavior to be used as baseline.

**Authors' static policies.** For what concerns the authors, we consider a template policy $\pi_A$ that is composed of two inner policies. The first one, that we denote by $\pi_{A,T}$, determines the decision of when stopping writing; the second one, that we denote by $\pi_{A,J}$, is triggered only when $\pi_{A,T}$ suggests to submit the paper and determines the decision of the journal to which the paper has to be submitted. For each one of $\pi_{A,T}$ and $\pi_{A,J}$ we consider a few variants, hence obtaining several variants of $\pi_A$ from their combinations.

Regarding $\pi_{A,T}$, we consider two variants:

- *Fixed* time (F). The author $a$ keeps writing the current paper $p' = p_{\text{draft}}(a)$ for a fixed amount $\hat{t}_{\text{fix}}$ of time steps; when $\hat{t}_{\text{fix}}$ is exceeded, $a$ submits $p'$ only if $p'$ got less then $\hat{n}_{\text{rej}}$ rejections; otherwise, $a$ starts a new paper—$\hat{t}_{\text{fix}}$ and $\hat{n}_{\text{rej}}$ are parameters of this policy. Hence, if $t(p') < \hat{t}_{\text{fix}}$ then $a$ action is Keep; otherwise, if $n_{\text{rej}}(p') < \hat{n}_{\text{rej}}$ the action is Submit-to-$j$ (with $j$ determined by $\pi_{A,J}$); otherwise the action is Restart.

- *Reasonable* time (R). The author $a$ keeps writing the current paper $p'$ until a satisfactory quality is obtained or a fixed amount of time $\hat{t}_{\text{max}}$ is exceeded; after having stopped the writing, the author behaves as in the Fixed case. We use the author's quality $q(a)$ as the satisfaction threshold for the paper quality; $\hat{t}_{\text{max}}$ and $\hat{n}_{\text{rej}}$ are parameters of this policy. Hence, if $t(p') < \hat{t}_{\text{max}} \wedge q(p') < q(a)$ then $a$ action is Keep; otherwise, if $n_{\text{rej}}(p') < \hat{n}_{\text{rej}}$ the action is Submit-to-$j$ (with $j$ determined by $\pi_{A,J}$); otherwise, the action is Restart.

Regarding the part of the policy $\pi_{A,J}$ that determines the journal to submit the paper to, we consider three variants:

- *Uniform* selection (U). The journal $j$ is chosen randomly with uniform probability in $J$.

- *Reasonable* selection (R). The author $a$, aware of the quality of their paper $p'$, selects a journal which is approximately on the same level. In practice, we first sort all the journals in ascending order according to their impact factor and group them in $n^{\text{R}}_{\text{bins}}$ bins. Then, we select the $i$-th bin, with $i = \max\left(0, \left\lceil \frac{q(p')}{q_{\max}} \left(n^{\text{R}}_{\text{bins}} - 1\right)\right\rceil - n_{\text{rej}}(p')\right)$. $n^{\text{R}}_{\text{bins}}$ is a parameter of this policy. Finally, we pick a journal $j$ by random choice with uniform probability inside the $i$-th bin. Intuitively, the author tailors their ambition based on the paper quality and lowers it upon each rejection. This policy, beyond mimicking a sound behavior, is consistent with the findings of Calcagno et al. (2012), who showed that authors are overall efficient in targeting the journal that would eventually publish their paper.

- *Eager* selection (E). The author selects the (approximately) best journal possible. In practice, we first partition the journals in $n^{\text{E}}_{\text{bins}}$ bins, as for the Reasonable case, and the select a random journal, with uniform probability, in the $i$-th bin, with $i = \max\left(0, n^{\text{E}}_{\text{bins}} - n_{\text{rej}}(p') - 1\right)$. $n^{\text{E}}_{\text{bins}}$ is a parameter of this policy.

Summing up, for the authors we consider six different static policies $\pi_A$, which are the different combinations of the variants for $\pi_{A,T}$ and $\pi_{A,J}$. We denote them as F+U, F+R, F+E, R+U, R+R, and R+E.

Downloaded from http://mitp.silverchair.com/isal/proceedings-pdf/isal2023/35/121/2355001/isal_a_00585.pdf by guest on 16 September 2024

**Journals static policies.** For the journals, which have to decide whether to accept or reject the submitted paper $p$, we define four different static policies $\pi_J$:

- *Accept all* (A). The journal $j$ always accepts $p$.
- *Mild* scrutiny (M). The journal $j$ accepts $p$ if $q(p) \geq q_{25\,\%}(j)$, where $q_{25\,\%}(j)$ is the first quartile of the quality of the papers $P_J(j)$ published by journal $j$. Otherwise, $j$ rejects $p$.
- *Strict* scrutiny (S). The journal $j$ accepts $p'$ if $q(p) \geq q_{50\,\%}(j)$. Otherwise, $j$ rejects $p$.
- *With citation* (C). The journal $j$ accepts $p$ if $R_P(p) \cap P(j) \neq \emptyset$, i.e., if $p$ references at least one paper published by the journal $j$.

## Model parameters

Simulating a realistic number of authors and journals would be computationally very challenging. For this reason, we decided to simulate $n_{\text{authors}} = 1000$ authors, as was previously done by Medo and Cimini (2016), and $n_{\text{journals}} = 60$ journals. We set the episode duration to $k_{\text{final}} = 240$ because we associate one time step $k$ to a month and we want our simulations to last 20 years, which is a reasonable length for a career in academia.

We set the remaining parameters as follows: $n_{\text{papers}} = 6000$, $q_{\text{max}} = 10$, $\delta q_{\text{min}} = 1$, $\delta q_{\text{max}} = 5$, $\alpha_q = 1$, $\beta_q = 1.5$, $\sigma_q = 1$, $\sigma_{\delta q} = 1$, $\gamma_{\text{ref}} = 5$, $y_{\text{max}} = 20$, $n_{\text{ref}} = 30$, $\hat{t}_{\text{fix}} = 4$, $\hat{t}_{\text{max}} = 12$, $\hat{n}_{\text{rej}} = 4$, $n^{\text{R}}_{\text{bins}} = 5$, and $n^{\text{E}}_{\text{bins}} = 3$.

## Validation simulations and results

For each combination of static policies, we performed 10 episodes with different random seeds, measured the global and local indexes, and averaged them across the episodes. Table 1 shows the results concerning the global indexes.

Several observations can be done based on Table 1. First, we see that, in general, when authors employ the Fixed policy for $\pi_{A,T}$, there is a greater number $|P|$ of published papers and a larger total quality $q_{\text{tot}}$ than with the Reasonable policy—we remark that the efficiency $e$ is linearly dependent on $q_{\text{tot}}$, so F+$*$ results in greater efficiency than R+$*$ too. Second, the Eager $\pi_{A,J}$ policy causes a lower number of papers and a lower total quality; an exception to this is when the author is eager and the journal requires a citation to accept the paper. In this case, the acceptance rate is higher since the authors submit their papers to the best journals, which will also probably be the journals from where they took their references. Third, it is interesting to note that when journals have a strict scrutiny, i.e., $\pi_J$ is Strict, even though the proportion of better quality papers grows with respect to the other $\pi_J$ policies (and equal authors' policy), the total quality produced is lower; i.e., $\rho_{q,\text{hi}}$ is greater with S than with M and A.

| $\pi_A$ | $\pi_J$ | $q_{\text{tot}}$ | $\lvert P\rvert$ | $e$ | $\rho_{q,\text{low}}$ | $\rho_{q,\text{mid}}$ | $\rho_{q,\text{hi}}$ |
|---|---|---|---|---|---|---|---|
| F+U | A | 247.6 | 65.0 | 40 | 47 | 37 | 16 |
| F+U | M | 200.6 | 40.2 | 32 | 25 | 50 | 25 |
| F+U | S | 157.5 | 27.8 | 25 | 15 | 50 | 35 |
| F+U | C | 161.2 | 41.9 | 26 | 46 | 38 | 16 |
| F+R | A | 247.3 | 65.0 | 40 | 47 | 37 | 16 |
| F+R | M | 222.9 | 51.4 | 36 | 38 | 43 | 19 |
| F+R | S | 202.7 | 42.6 | 32 | 31 | 47 | 22 |
| F+R | C | 140.6 | 31.2 | 23 | 36 | 41 | 23 |
| F+E | A | 246.2 | 65.0 | 39 | 47 | 37 | 16 |
| F+E | M | 173.7 | 33.7 | 28 | 23 | 51 | 26 |
| F+E | S | 155.4 | 28.9 | 25 | 17 | 53 | 30 |
| F+E | C | 177.6 | 46.9 | 28 | 47 | 37 | 16 |
| R+U | A | 148.2 | 39.2 | 24 | 49 | 34 | 17 |
| R+U | M | 130.1 | 27.5 | 21 | 32 | 44 | 24 |
| R+U | S | 116.3 | 21.7 | 19 | 21 | 47 | 32 |
| R+U | C | 108.7 | 28.5 | 17 | 48 | 35 | 17 |
| R+R | A | 149.4 | 39.2 | 24 | 49 | 33 | 18 |
| R+R | M | 141.9 | 33.2 | 23 | 41 | 38 | 21 |
| R+R | S | 134.2 | 29.9 | 22 | 37 | 41 | 22 |
| R+R | C | 97.2 | 21.0 | 16 | 36 | 37 | 27 |
| R+E | A | 150.0 | 39.2 | 24 | 48 | 34 | 18 |
| R+E | M | 119.2 | 23.7 | 19 | 28 | 45 | 27 |
| R+E | S | 114.6 | 22.0 | 18 | 22 | 49 | 29 |
| R+E | C | 119.5 | 31.3 | 19 | 48 | 34 | 18 |

Table 1: Results with the static policies. $q_{\text{tot}}$ and $|P|$ are in thousands; $e$, $\rho_{q,\text{low}}$, $\rho_{q,\text{mid}}$, and $\rho_{q,\text{hi}}$ are percentages.

While the above considerations are qualitatively sound, they cannot be directly used for validating our system with respect to the reality because in the latter the global indexes are not available. For this reason, we also measured the local indexes, starting from the h-index.

Figure 1 shows the relation between h-index and author's quality, in the form of scatter plots, for each first episode (i.e., the episode with random seed 0) of each $\pi_A$ policy, coupled with the Mild policy for the journals.

From the figure, we can see that the h-index is in general well correlated with the author's quality, a finding that is consistent with the literature on bibliometrics (Van Raan, 2006). Moreover, the authors' h-index is impacted more by $\pi_{A,T}$ than by $\pi_{A,J}$, i.e., more by when to submit than by where to submit to. We can also note that with the Reasonable journal selection there are no authors with h-index equal to 0. This happens because authors choose more quality-appropriate journals to send their papers to, hence also the worst authors are able to publish some papers (on the worst journals).

To attempt to perform a quantitative validation of our model with real data, we compared the distributions of our h-index against the real one. For the latter, we collected

Downloaded from http://mitp.silverchair.com/isal/proceedings-pdf/isal2023/35/121/2355001/isal_a_00585.pdf by guest on 16 September 2024

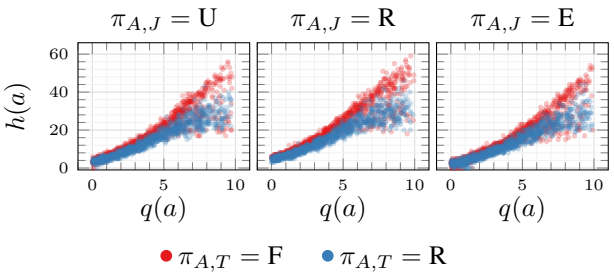


Figure 1: Authors' h-index $h(a)$ vs. quality $q(a)$ for six static author policies and the Mild journal policy, one point for each author of the first episode.

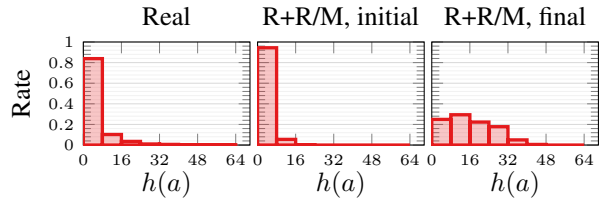


Figure 2: Distribution of the h-index $h(a)$ for real data and simulation (with R+R/M at $k = 0$ and $k = k_{\text{final}}$).

real data querying the Semantics Scholar API for the h-index of 3000 random authors, selecting the first 1000 results for each of the following queries for the author's last name: Smith, Gupta, and Li. Figure 2 compares the real distribution against two simulated ones for the R+R/M policy combination: at the beginning of episode and at the end. We chose the R+R/M policy combination because it appears to be the more realistic one: however, we verified that the distributions for most of the other policy combinations were qualitatively similar.

From Figure 2, we can see that our system starts with a h-index distribution which is very similar to the real one, and then flattens. This can be explained by the fact that in our model (a) all authors employ the same policy and (b) all authors have the same career duration. We speculate that, in the real data, a vast majority of the authors are in the initial stage of their career.

We performed a similar quantitative validation for the journals impact factor. We collected journals bibliometric data relative to the year 2018 obtained from Scopus and compared them to simulated data, in terms of distribution. Figure 3 shows the results of this analysis, similarly to Figure 2.

From Figure 3, we can see that throughout the simulation our system maintains a distribution comparable to the real one, validating our model.

Finally, we analyzed the relation between the quality of a journal $j$, measured as the median quality $\overline{q}(j)$ of the papers published in $j$ at the end of the episode, and the impact factor

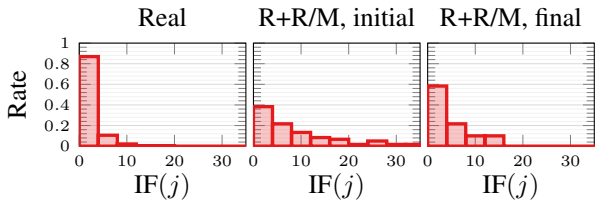


Figure 3: Distribution of the h-index IF($j$) for real data and simulation (with R+R/M at $k = 0$ and $k = k_{\text{final}}$).

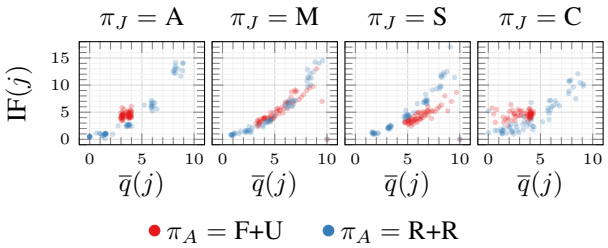


Figure 4: Journals impact factor IF($j$) vs. quality $\overline{q}(j)$ for eight static policy combinations, one point for each journal of the first episode.

IF($j$) of $j$. Figure 4 shows the results of this analysis in the form of four scatter plots, once for each of the journal policies: for each $\pi_J$, the color of the markers encode one of the two authors' policies F+U and R+R.

From Figure 4, we can see that when the journals have the All policy and the authors have a F+U policy, the impact factor is not representative of the quality of a journal. This is what we would have expected, since all journals publish whatever paper they receive and the authors behave randomly. Both the Mild and Strict policies, results in a good correlation between journals quality and impact factor, with Mild appearing very robust with respect to the authors' behavior. Finally, the Citation policy appears to favor a correlation between impact factor and quality only if the authors behave accordingly to reasonable choices.

## Incentives and their impact

The results we obtained with the static policies are interesting, but our main focus is on the effects of authors' incentives on the scholarly publishing system. In order to investigate this, we need to make authors behave driven by an externally dictated incentive. To this end, we *learn* the authors' policy with reinforcement learning (RL), instead of using a static one, and use a reward that encodes the incentive when learning.

In particular, we define an RL problem where the goal is to maximize a given reward function. We learn an authors' policy using an RL technique. Finally, we use the learned policy for performing a number of simulations where the leaning does not occur anymore and measure the global and

Downloaded from http://mitp.silverchair.com/isal/proceedings-pdf/isal2023/35/121/2355001/isal_a_00585.pdf by guest on 16 September 2024

local indexes as we did for the static policies. We repeat this procedure with five different reward functions representing different incentives.

## Learnable template policy

We define a learnable template policy $\ell$ for the author that is based on the idea of author's current *ambition* $m \in \{1, \dots, m_{\text{max}}\}$. We embed an RL agent inside the author.

At each time step, the RL agent performs one among the following five actions: keep writing and decrease ambition (*Keep-$m$-*), keep writing and leave ambition untouched (*Keep-$m$=*), keep writing and increase ambition (*Keep-$m$+*), *Restart* writing, and *Submit* the current paper. The first three actions of the RL agent trigger a Keep action of the author agent and modify the value of $m$ accordingly (clipping it to its domain); the Restart action triggers a Restart action; the Submit action triggers a Submit-to-$j$ action where the journal $j$ is selected similarly to the Reasonable $\pi_{A,J}$ policy, but based on the ambition $m$ for binning and choosing the bin: we partition the journals in $m_{\text{max}}$ bins and take one random journal in the $m$-th bin, i.e., the bin corresponding to the author's current ambition.

As RL agent observation, we use the tuple $(m, \tau_q, \tau_{\text{rej}}, t_{\text{quart}})$, where $m$ is the current ambition, $\tau_q \in \{0, 1\}$ is a binary indication about the current paper quality compared to the author's quality, i.e., $\tau_q = \mathbb{1}(q(p') \geq q(a))$, $\tau_{\text{rej}} \in \{0, 1\}$ is a binary indication about the paper being already been rejected, i.e., $\tau_{\text{rej}} = \mathbb{1}(n_{\text{rej}}(p') \geq 1)$, and $t_{\text{quart}} \in \{0, 1, 2\}$ is a ternary indication on the time spent on the current paper, with $t_{\text{quart}} = \min\left(2, \left\lfloor \frac{1}{4} t(p') \right\rfloor\right)$.

Summarizing, $\ell$ takes actions based the internal RL agent actions and on $m$, which represents the current author ambition. Internally, $m$ is modified by the RL agent that takes one among five actions based on one among $m_{\text{max}}(2 \cdot 2 \cdot 3) = 12 m_{\text{max}}$ possible observations.

In our experiments, we initialize the author's ambition at $k = 0$ to $m = \left\lfloor \frac{1}{2} m_{\text{max}} \right\rfloor$ and we set $m_{\text{max}} = 5$.

## Incentive-encoding reward

We experiment with the following reward variants.

The reward is a number $r \in \mathbb{R}$ made available to the RL agent embedded in $\ell$ at each time step. In all cases, we set $r = 0$ in the time step following a Keep-$m$* or Restart author's action. The reward variants differ in the value they get upon a Submit action:

- *Acceptance/rejection* reward ($\text{AR}_x$). We set $r = 1$ upon acceptance and $r = -x$ upon rejection. We experimented with $x \in \{0, \frac{1}{2}, 1\}$.
- *Quality* reward (Q). We set $r = q(p)$ upon acceptance of the submitted paper $p$ and $r = 0$ upon rejection.
- *Impact factor* reward (IF). We set $r = \text{IF}(j)$ upon acceptance by a journal $j$ and $r = 0$ upon rejection.

Intuitively the three reward variants represent, respectively, the case ($\text{AR}_x$) in which an author is happy or sad just for the journal response, regardless of the paper and journal quality, the case (Q) in which an author is happy with acceptance proportionally to the quality of the accepted paper, and the case (IF) in which an author is happy with acceptance proportionally to the impact factor of the accepting journal.

We denote the policies learned with these rewards as $\ell\text{AR}_0$, $\ell\text{AR}_{\frac{1}{2}}$, $\ell\text{AR}_1$, $\ell$Q, and $\ell$IF.

## Learning the learnable policies

We stated the RL problem with discrete action and observation spaces. Hence, we use the Q-learning RL algorithm (Watkins, 1989), which is a natural choice for this kind of problems.

In brief and intuitively, Q-learning expresses the policy in a tabular form, with rows corresponding to actions, columns corresponding to observations, and cells containing a value that says how convenient is to perform an action (cell row) given an observation (cell column). Whenever the Q-learning agent has to take an action given an observation, it simply selects the one of the observation column with the largest cell value or, only while learning and with a small decaying probability $\epsilon$, a random action—this is done to balance between exploration and exploitation. Moreover, while learning, values in the cells are increased or decreased based on the reward obtained at the next time step. We refer the reader to (Watkins, 1989) for more details.

In our work, we use Q-learning in a multi-agent system. That is, several agents share the same policy and, at each time step, receive different observations, different rewards, and take different actions. We exploit the availability of many triplets (observation, reward, action) at each time step for learning the policy faster. In particular, after each time step, for each cell of the tabular policy we average the modifications resulting from all the triplets collected by the agents and apply one single modification that will impact the policy at the next time step.

For each of the five rewards described above, we learned 5 learnable policies by performing 200 consecutive policies with a reduced size model with 500 authors (instead of 1000) and the Mild policy for the journals—we chose this policy as it proved to be the most robust with respect to authors' behavior, as shown in Figure 4. We hence obtained 25 learnable author policies: for each of them, we performed other simulations keeping the policy steady, i.e., without learning. We set Q-learning as follows: decaying probability $\epsilon = \frac{1}{j}$, learning rate $\lambda = \frac{1}{j+1}$, and discount rate $\gamma = 0.95$, where $j \in \{1, \dots, 200\}$ is the index of the episode.

## Results

Like we did for the static policies, for each one of the optimized learnable policies, we performed 10 episodes with 1000 authors. This means that, for each reward function, we

Downloaded from http://mitp.silverchair.com/isal/proceedings-pdf/isal2023/35/121/2355001/isal_a_00585.pdf by guest on 16 September 2024

| $\pi_A$ | $q_{\text{tot}}$ | $\lvert P\rvert$ | $e$ | $\rho_{q,\text{low}}$ | $\rho_{q,\text{mid}}$ | $\rho_{q,\text{hi}}$ |
|---|---|---|---|---|---|---|
| $\ell\text{AR}_0$ | 150.4 | 42.9 | 24 | 49 | 48 | 3 |
| $\ell\text{AR}_{\frac{1}{2}}$ | 232.9 | 67.1 | 37 | 49 | 49 | 2 |
| $\ell\text{AR}_1$ | 160.0 | 34.8 | 26 | 32 | 49 | 19 |
| $\ell$Q | 198.1 | 56.7 | 32 | 48 | 49 | 3 |
| $\ell$IF | 70.1 | 17.4 | 11 | 38 | 50 | 12 |

Table 2: Results with the learnable policies. $q_{\text{tot}}$ and $|P|$ are in thousands; $e$, $\rho_{q,\text{low}}$, $\rho_{q,\text{mid}}$, and $\rho_{q,\text{hi}}$ are percentages.

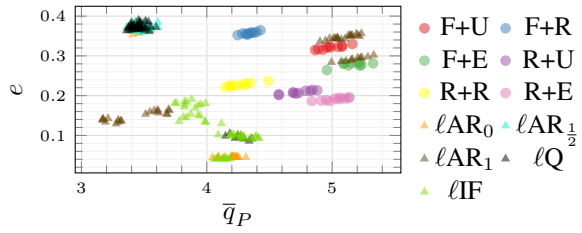


Figure 5: System efficiency $e$ vs. paper mean quality $\overline{q}_P$ for all the episodes with all $\pi_A$ and $\pi_J = \text{M}$.

averaged the results of 50 simulations; 10 for each one of the 5 learned policies. Table 2 shows the results in terms of the global indexes.

We note that the $\ell\text{AR}_{\frac{1}{2}}$ policy is the one which leads to the highest amount of science produced, i.e., $q_{\text{tot}}$, and hence the greatest system efficiency $e$: on the other hand, $\ell\text{AR}_{\frac{1}{2}}$ results in the lowest portion of papers of high quality. Increasing the penalty for the rejections with the $\ell\text{AR}_1$ policy, we see that the number of published papers decreases, but the portion of better quality publications is the highest. For what concerns $\ell$Q, we can see that the authors prefer to publish many lower or median quality papers, rather than publishing a few good quality works; the same holds for $\ell\text{AR}_0$, which is not surprising, since both cases reward acceptance and do not penalize rejections. Finally, we can see that when the authors are incentivized to publish on the most prestigious journals, i.e., with $\ell$IF, the lowest number of papers and the lowest total quality is produced, with a huge waste of scientific production.

Figure 5 shows the results obtained with the static and learnable author policies (and the Mild journal policy) in a disaggregated form, with a scatterplot having one point per simulation. In particular, it shows the efficiency $e$ vs. the paper mean quality $\overline{q}_P = \frac{q_{\text{tot}}}{|P|}$.

The figure makes apparent the trade-off between overall efficiency and mean paper quality. Some of the author policies place on the Pareto frontier, including $\ell\text{AR}_1$, F+R, and others. The figure also shows that not all the policies learned with the same behavior actually resulted in the same global indexes: e.g., some of the five $\ell\text{AR}_1$ policies correspond to simulations not being on the $\overline{q}_P, e$ Pareto frontier.

| $\pi_A$ | $\overline{h}_{\text{low}}$ | $\overline{h}_{\text{mid}}$ | $\overline{h}_{\text{hi}}$ |
|---|---|---|---|
| F+U | 15 | 18 | 20 |
| F+R | 18 | 21 | 23 |
| F+E | 13 | 17 | 18 |
| R+U | 13 | 14 | 15 |
| R+R | 14 | 16 | 18 |
| R+E | 11 | 13 | 14 |

| $\pi_A$ | $\overline{h}_{\text{low}}$ | $\overline{h}_{\text{mid}}$ | $\overline{h}_{\text{hi}}$ |
|---|---|---|---|
| $\ell\text{AR}_0$ | 8 | 13 | 18 |
| $\ell\text{AR}_{\frac{1}{2}}$ | 13 | 20 | 28 |
| $\ell\text{AR}_1$ | 11 | 14 | 18 |
| $\ell$Q | 11 | 17 | 25 |
| $\ell$IF | 3 | 7 | 11 |

Table 3: Mean h-index $\overline{h}$ for the three categories of authors based on their productivity tertile, with all $\pi_A$.

Finally, we analyzed how the authors' productivity $\delta q$ impacts their h-index with different author policies. To this end, we divided the authors in tertiles based on their $\delta q$ and computed the mean h-index of each tertile ($\overline{h}_{\text{low}}$, $\overline{h}_{\text{mid}}$, and $\overline{h}_{\text{hi}}$). Table 3 shows the results of this analysis for a all the static and learnable author policies.

Table 3 shows that, in all cases, there is a positive correlation between the authors' h-index and their productivity. However, the learnable policies tend to make the difference between least and most productive authors more apparent, in particular $\ell$IF. Since authors' quality and productivity are, in our model, independent, this means that incentives tend to make more productive, rather than better, authors stand out.

## Concluding remarks

We proposed a multi-agent model of the scholarly publishing system and we used it for studying, through simulation, the overall impact of author's incentives on the output of the system. We used reinforcement learning for modeling agents whose behavior is driven by an incentive, encoding the incentive in the reward function.

We validated our model through several experiments with different hand-written agent policies and found that it approximates the dynamics of the real counterpart. We then experimented with different incentives (including those representing personal satisfaction upon paper acceptance, quality of one's paper, or reputation of the journal) and obtained interesting findings. Namely, we found that (a) regardless of incentives, quantity tends to be favored over quality and (b) when the journal reputation is the incentive, there is a waste of authors' effort.

While our model cannot capture all the many facets of scholarly publishing, we believe that it proves that the overall behavior of human communities can be approximated by multi-agent systems.

## References

Abalkina, A. (2021). Detecting a network of hijacked journals by its archive. *Scientometrics*, 126:7123 – 7148.

Baccini, A., De Nicolao, G., and Petrovich, E. (2019). Citation

Downloaded from http://mitp.silverchair.com/isal/proceedings-pdf/isal2023/35/121/2355001/isal_a_00585.pdf by guest on 16 September 2024

gaming induced by bibliometric evaluation: A country-level comparative analysis. *PLoS One*, 14(9):e0221212.

Barnes, C. (2017). The h-index debate: an introduction for librarians. *The Journal of Academic Librarianship*, 43(6):487–494.

Bartoli, A., De Lorenzo, A., Medvet, E., and Tarlao, F. (2016). Your paper has been accepted, rejected, or whatever: Automatic generation of scientific paper reviews. In *Availability, Reliability, and Security in Information Systems*, pages 19–28, Cham. Springer International Publishing.

Bartoli, A. and Medvet, E. (2014). Bibliometric evaluation of researchers in the internet age. *The Information Society*, 30(5):349–354.

Bartoli, A. and Medvet, E. (2020). Exploring the Potential of GPT-2 for Generating Fake Reviews of Research Papers. In *Fuzzy Systems and Data Mining VI*, pages 390–396. IOS Press.

Bornmann, L. and Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, 66(11):2215–2222.

Byrne, J. A., Grima, N., Capes-Davis, A., and Labbé, C. (2019). The possibility of systematic research fraud targeting understudied human genes: Causes, consequences, and potential solutions. *Biomarker Insights*, 14.

Cabanac, G., Labbé, C., and Magazinov, A. (2021). Tortured phrases: A dubious writing style emerging in science. evidence of critical issues affecting established journals.

Calcagno, V., Demoinet, E., Gollner, K., Guidi, L., Ruths, D., and de Mazancourt, C. (2012). Flows of research manuscripts among scientific journals reveal hidden submission patterns. *Science*, 338(6110):1065–9.

Corrall, S., Kennan, M. A., and Afzal, W. (2013). Bibliometrics and research data management services: Emerging trends in library support for research. *Library trends*, 61(3):636–674.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. In *Science*, volume 178(4060), pages 471–479.

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., and Rafols, I. (2015). Bibliometrics: the leiden manifesto for research metrics. *Nature*, 520(7548):429–431.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. In *Proceedings of the National academy of Sciences*, volume 102(46), pages 16569–16572.

Medo, M. and Cimini, G. (2016). Model-based evaluation of scientific impact indicators. *Physical Review E*, 94(3):032312.

Peroni, S., Ciancarini, P., Gangemi, A., Nuzzolese, A. G., Poggi, F., and Presutti, V. (2020). The practice of self-citations: a longitudinal study. *Scientometrics*, pages 1–30.

Rahal, R., Fiedler, S., and Adetula, A., e. a. (2023). Quality research needs good working conditions. *Nature Human Behaviour*, 7:164–167.

Ruocco, G., Daraio, C., Folli, V., and Leonetti, M. (2017). Bibliometric indicators: the origin of their log-normal distribution and why they are not a reliable proxy for an individual scholar's talent. *Palgrave Communications*, 3:17064.

Smaldino, P. E. and McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, 3(160384).

Van Raan, A. F. (2006). Comparison of the hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *scientometrics*, 67(3):491–502.

Watkins, C. J. (1989). *Learning from Delayed Rewards*. PhD thesis, King's College, Cambridge, UK.

Downloaded from http://mitp.silverchair.com/isal/proceedings-pdf/isal2023/35/121/2355001/isal_a_00585.pdf by guest on 16 September 2024