Rating and Matching in Peer Review Systems

Yuanzhang Xiao, Florian Dörfler, and Mihaela van der Schaar

Abstract— Peer review (e.g., review of research papers) is essential for the success of the scientific community. In peer review, the reviewers voluntarily exert *costly* effort in reviewing papers. Hence, it is important to design mechanisms to elicit high effort from reviewers. Exploiting the fact that the researchers interact with each other repeatedly (e.g., by submitting and reviewing papers over years), we propose a rating and matching mechanism to elicit high effort from reviewers. Our proposed mechanism overcomes two major difficulties, namely adverse selection (i.e., the unidentifiable quality of heterogeneous reviewers) and moral hazard (i.e., the unobservable effort levels from reviewers). Specifically, our proposed mechanism assigns and updates ratings for the researchers, and matches researchers' papers to reviewers with similar ratings. In this way, the mechanism identifies different types of reviewers by their ratings, and incentivizes different reviewers to exert high effort.

Focusing on the matching rule, we first provide design guidelines for a general matching rule that leads the system to an equilibrium, where the reviewers' types are identified and their high efforts are elicited. Then we study in detail a baseline matching rule that assigns each researcher's paper to one of the two reviewers with the closest ratings, provide guidelines of how to choose the initial ratings, and analyze equilibrium review quality and equilibrium ratings. Finally, we extend the baseline matching rule to two classes. The first extension provides extra reward and/or punishment by adjusting the probabilities of matching each researcher's paper to its neighbors. The second extension provides extra reward and/or punishment by allowing to match each researcher's paper to reviewers other than its neighbors. We prove that it is beneficial (in the sense that the optimal equilibrium review quality is higher) to reward reviewers in the first extension, and to punish reviewers in the second extension, due to the different ways the reward and punishment are carried out. We also prove that our proposed matching rules elicit much higher effort from reviewers, compared to matching rules that mimic the current mechanisms of assigning papers.

I. Introduction

Peer review (e.g., review of research papers) is an integral and essential part of the academia. The efficiency of the peer review system has a critical impact on the quality of a research community. A key problem in the peer review system is that the reviewers voluntarily exert effort in reviewing papers, and that the effort are costly for the reviewers. Hence, it is important to design mechanisms in which the reviewers have incentives to review papers with high effort, such that the efficiency of the peer review system is improved.

Based on the fact that the researchers interact with each other repeatedly (e.g., by submitting and reviewing papers

Xiao and van der Schaar are with Department of Electrical Engineering, UCLA. Emails: `yxiao@seas.ucla.edu` and `mihaela@ee.ucla.edu`

Dörfler is with Automatic Control Laboratory, ETH Zurich, Switzerland. Email: `dorfler@ethz.ch`

over time), we propose a rating and matching mechanism to elicit high effort from the reviewers. The basic idea of our proposed mechanism is to assign a rating for each researcher and update their ratings based on their past review quality reported by the authors. Next we match the papers from high-rating researchers more likely to high-rating reviewers. Hence, the researchers have incentives to exert high effort levels to improve their ratings and the chances of getting high-quality reviewers.

There are two major challenges in designing an efficient rating and matching mechanism. First, even though the quality of a review can be assessed by the authors, there is no way to observe the effort level of the reviewer. This is because the quality of the review is an unknown function of the effort level exerted by the reviewer. The problem of unobservable effort levels is called *moral hazard* problems in economics literature. Second, since the reviewers are heterogeneous, their review quality can be different even if they exert the same amount of effort. This problem of unidentifiable quality (i.e., the mapping from a reviewer's effort level to the review quality) of a reviewer is called *adverse selection* problems in economics. In the presence of both moral hazard and adverse selection problems, it is difficult to identify high-quality and low-quality reviewers, and the effort levels exerted by them.

In this paper, we design the rating update rule and the matching rule, such that we can identify different types of reviewers and incentivize them to choose high effort levels in the equilibrium. The rating update rule ensures that the ratings of the reviewers truly reflect their review quality. The matching rule incentivizes different types of reviewers to exert high effort. In the equilibrium of the system, more capable reviewers with higher quality and lower cost in reviewing provide higher-quality review, and hence have higher ratings which in turn confirm their quality.

We first propose a baseline matching rule, which simply matches a researcher's paper to reviewers with closest (higher and lower) ratings. The probability of being matched to a higher-rating or lower-rating reviewer depends on the distance from the researcher's rating to the reviewers' ratings. We provide design guidelines of how to choose the initial ratings, prove the convergence to the equilibrium, and analyze the equilibrium review quality and rating of researchers with different quality. We then extend the baseline matching rule in two directions. Both extensions enable us to construct a class of matching rules and to tune the extent to which the matching rules reward and/or punish reviewers. The first extension provides extra reward and/or punishment by adjusting the probabilities of matching each

researcher's paper to its neighbors. The second extension provides extra reward and/or punishment by allowing to match each researcher's paper to reviewers other than its neighbors. Our interesting finding shows that it is beneficial (in the sense that the optimal equilibrium review quality is higher) to reward reviewers in the first extension, and to punish reviewers in the second extension, due to the different ways the reward and punishment are carried out. Our results show the importance of designing the correct reward and punishment mechanisms for different types of matching rules.

II. RELATED WORKS

To the best of our knowledge, no prior work has proposed joint matching and rating mechanisms for moral hazard and adverse selection problems. Next, we discuss related works on rating mechanisms and those on matching separately.

A. Works on Rating Mechanisms

1) Rating Mechanisms for Review Score Aggregation: Some works ([1] and references therein) on rating mechanisms study how to aggregate review scores (i.e., ratings) to ensure the correct decisions on the paper/proposal acceptance. The ratings in [1] are provided by the *reviewers* to evaluate *papers*, while the ratings in our work are maintained by the *editorial system* to evaluate *reviewers*. Moreover, the focus in [1] is to make the correct acceptance/rejection decision by aggregating the ratings from the reviewers, while the focus here is to incentivize the reviewers to exert high effort levels. There is no notion of "effort" in [1].

2) Rating Mechanisms for Effort Elicitation: Some works on rating mechanisms study how to elicit high effort through ratings. See our recent works applied to peer-to-peer systems [3][4], which are based on the seminal work on social norms [2]. The problem studied in these works [2]–[4] is similar to the one studied in this work, namely how to design rating mechanisms to incentivize agents to exert costly effort.

However, the works [2]–[4] focus exclusively on the moral hazard problem (i.e., the problem of elicit unobservable effort), and ignore the adverse selection problem (i.e., the problem of learning the heterogeneous review quality) by assuming *homogeneous* agents. In this work, we assume *heterogeneous* agents and deal with both the moral hazard and adverse selection problems. In [2]–[4], due to the homogeneity of agents, binary ratings are usually sufficient to identify whether a player has behaved well or badly. In contrast, in this work, the rating is continuous, such that the rating mechanism can identify not only whether a player has behaved well or badly, but also its review quality.

Another key difference is that we extensively study and design the matching rule, while a fixed uniformly random matching rule is used in [2]–[4]. In our setting, we will prove that a uniformly random matching rule will result in the lowest review quality in the equilibrium.

B. Works on Matching

1) Matching in Resource Allocation and Exchange: There is a huge literature on matching in resource allocation (e.g., allocation of schools to students [6]) and exchange (e.g., kidney exchange [7]). To some extent, they focus on adverse selection problems (i.e., identifying the benefit from matching to a particular agent), but do not consider moral hazard problems (i.e., eliciting high effort from agents). In particular, they do not have the notation of "effort" at all. Once an agent (e.g., a student) is matched to another (e.g., a school), the benefit (obtained by this student) is fixed. In contrast, in our paper, once a paper is matched to a reviewer, the benefit from the review depends not only on the reviewer's quality, but also on its effort. The goal of our paper is to incentivize the reviewers to exert high effort, which is different from the goal of the works on matching in resource allocation and exchange.

2) Matching in Repeated Games: Repeated games research in economics assumes that agents interact either with the same agents all the time (direct reciprocity), or with anonymous agents randomly (indirect reciprocity) [2], as is the case in our work. However, the existing research on indirect reciprocity assumes that the agents are *uniformly* randomly matched according to *exogenous* matching rules. Our work significantly differs from all this literature: the proposed matching rules are *endogenous* and directly affect agents' incentives. This makes our study both significantly different and much more challenging than existing works.

III. MODEL

A. Basic Setup

Consider a community $\mathcal{N} = \{1, \ldots, N\}$ of N researchers. Each researcher acts both as an author, who benefits from reviews of its submitted paper, and as a reviewer, who exerts effort in reviewing papers. A designer (e.g., the editor of a journal) aims to design an editorial policy that incentivizes the researchers to exert high effort in reviewing papers. The editorial policy includes two parts: the rating mechanism that assigns and updates a rating $\theta_i \in \mathbb{R}_+$ for each researcher i, and the matching rule that matches papers with reviewers (possibly based on the ratings). In the following, we write the rating profile, namely the ratings of every researcher, as $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)$. The rating profile is known only to the designer. We define the rating distribution $d(\boldsymbol{\theta})$ as a vector of length N that is an ordered list of all the ratings (from high ratings to low ratings). We denote the jth element of the rating distribution by $d(\boldsymbol{\theta})_j$. Notice that $d(\boldsymbol{\theta})_j$ is not necessarily identical to θ_j. Hence, the rating distribution does not disclose any information about the identities of the researchers.

Time is slotted as $t = 1, 2, \ldots$. In each time slot t, the entities in the system moves in the following order:[1]

1) The designer publishes the **rating distribution** $d(\boldsymbol{\theta})$, and informs each researcher i of its rating θ_i and its ranking k_i (i.e., ordered position) in the rating distribution.

---

[1]Throughout the paper, the superscript $(\cdot)'$ on a function refers to the derivative, and the superscript $(\cdot)^t$ refers to the variable under consideration at time point $t \in \mathbb{Z}_+$.

2) Each researcher submits a paper.
3) The designer matches each researcher i's paper to another reviewer (i.e., one of the other researchers) according to a probabilistic **matching rule** $m_{k_i j}$: $(d(\boldsymbol{\theta})_{k_i}, d(\boldsymbol{\theta})_j) \mapsto [0, 1]$. The matching rule $m_{k_i j}$ determines the probability $m_{k_i j}(d(\boldsymbol{\theta})_{k_i}, d(\boldsymbol{\theta})_j)$ that the paper of the researcher with the k_ith highest rating is matched to the reviewer with the jth highest rating. From the definition, we can see that the matching rule depends only on the researchers' rankings in their ratings, but not on their identities.
4) Each researcher j chooses an **effort level** $e_j^t \in [0, e_j^{\max}]$ to review the paper, where E_j is j's maximum effort level. The **quality** of researcher j's review then depends on its effort as $q_j(e_j^t)$, where $q_j : \mathbb{R}_+ \to \mathbb{R}_+$ is the review quality function.
5) Each researcher i receives **benefit** $b_i(q_j(e_j^t))$ from the obtained review, where $b_i : \mathbb{R}_+ \to \mathbb{R}_+$ is i's benefit function, and incurs a **cost** of $c_i(e_i^t)$ for reviewing a paper, where $c_i : \mathbb{R}_+ \to \mathbb{R}_+$ is i's cost function.
6) Each researcher i sends a **report** r_i^t about the review to the designer. We assume that the report accurately reflects the reviewer's review quality, namely $r_i^t = q_j(e_j^t)$.
7) The designer updates the researchers' ratings. We write the rating update rule as $\pi : (\theta_j^t, r_i^t) \mapsto \theta_j^{t+1}$. For fairness, the **rating update rule** is identical for all reviewers and is given by a convex combination of the reviewer's old rating and the report about its review quality:[2]

$$\pi(\theta_j^t, r_i^t) = \begin{cases} (1 - \mu) \cdot \theta_j^t + \mu \cdot r_i^t, & j \text{ reviewed papers,} \\ \theta_j^t, & \text{otherwise.} \end{cases} \quad (1)$$

where $\mu \in (0, 1)$ is a constant step size.

Note that each researcher i has a maximum effort level $e_i^{\max}$, and hence has a maximum review quality $q_i(e_i^{\max})$. Since the new rating is the convex combination of the old rating and the review quality, given any initial rating θ_i^0, researcher i's rating can only be in the interval $[0, \theta_i^{\max}]$, where $\theta_i^{\max} \triangleq \max\{\theta_i^0, q_i(e_i^{\max})\}$. In other words, the possible ratings of each researcher are contained in the compact set $[0, \theta_i^{\max}]$.

Throughout the paper, we make the following assumption about the cost function and the review quality function.

Assumption 1 (Cost, Review Quality, and Benefit): Each researcher i's cost function $c_i(\cdot)$, review quality function $q_i(\cdot)$, and benefit function $b_i(\cdot)$ satisfy the following:

- The cost $c_i(\cdot)$ is strictly convex, strictly increasing, and differentiable in the effort level e_i.
- The review quality $q_i(\cdot)$ is concave, increasing. and differentiable in the effort level e_i. In addition, $q_i'(0)$ is bounded.

- The benefit $b_i(\cdot)$ is increasing, concave, and differentiable in review quality q_j.
- Without loss of generality, we normalize $c_i(0) = 0$, $q_i(0) = 0$, and $b_i(0) = 0$.

B. Information – Who Knows What

We summarize the information known to each entity in the system.

1) The designer: The designer receives reports r_i of review quality, and keeps the rating θ_i^t for each researcher i at each time slot t. Hence, the designer knows the identity of the researcher at the k_ith position of the rating distribution. However, it does not know the review quality functions $q_i(\cdot)$, the benefit functions $b_i(\cdot)$, or the cost functions $c_i(\cdot)$.

2) Each researcher i: Each researcher i knows its own review quality function $q_i(\cdot)$, benefit function $b_i(\cdot)$, and cost function $c_i(\cdot)$, but does not know the above functions of the other researchers. It also knows the matching rule m and the rating update rule π. It knows its own rating θ_i^t at each time slot t, the rating distribution $d(\boldsymbol{\theta})^t$ at each time slot t, and its position in the rating distribution k_i^t. Note that each researcher i does not know the ratings of other researchers, or the identity of its reviewer.

C. Payoffs and Equilibrium

In each time slot t, researcher i's expected payoff is its expected benefit from the review of its paper minus the expected cost of reviewing another researchers' papers. We write researcher i's expected payoff as $u_i(m, \theta_i, d, \boldsymbol{e})$, which depends on the matching rule m, its own rating θ_i, the rating distribution d, and all the researchers' effort levels $\boldsymbol{e} \triangleq (e_1, \ldots, e_N)$. The *expected payoff* is defined as

$$u_i(m, \theta_i, d(\boldsymbol{\theta}), \boldsymbol{e}) = \sum_{j \neq k_i} m_{k_i j}(d(\boldsymbol{\theta})_{k_i}, d(\boldsymbol{\theta})_j) \cdot b_i(q_j(e_j))$$
$$- \left[\sum_{j \neq k_i} m_{j k_i}(d(\boldsymbol{\theta})_j, d(\boldsymbol{\theta})_{k_i}) \right] \cdot c_i(e_i). \quad (2)$$

The goal of each researcher i is to choose an optimal sequence of effort levels $\{e_i^{t,*}\}_{t=0}^{\infty}$ to maximize the *discounted average of expected payoffs*, namely

$$\{e_i^{t,*}\}_{t=0}^{\infty} = \arg \max_{\{e_i^t \in \mathbb{R}_+\}_{t=0}^{\infty}}$$
$$\mathbb{E}\left\{ (1 - \delta_i) \sum_{t=0}^{\infty} \delta_i^t u_i(m, \theta_i^t, d^t, e_i^t, \boldsymbol{e}_{-i}^t) \right\}, \quad (3)$$

where $\boldsymbol{e}_{-i}^t$ is the effort levels chosen by all the researchers other than i, and $\delta_i \in [0, 1)$ is researcher i's *discount factor*. A researcher's discount factor reflects its patience. We take expectation $\mathbb{E}\{\cdot\}$ because the rating update is random, namely a researcher's rating is either updated or kept the same depending whether it is matched to a paper.

Note that the optimization problem (3) is very hard, if not impossible, to solve. The difficulty lies in the coupling of one researcher's decisions over time and the coupling of different researchers' decisions. First, the researcher's current

---

[2]Under the assumption that j's review quality $q_j(e_j)$ is perfectly observed, it does not matter how many reports about j's review quality are received.

decision (i.e., effort level) affects not only its current payoff (through the cost), but also its future ratings and hence future payoffs. Second, the researcher's payoff is affected by the others' decisions (through the benefit). However, since a researcher has no knowledge about the others' review quality functions, benefit functions, or cost functions, it cannot predict the others' effort levels and the evolution of the rating distributions. In summary, a researcher cannot solve the optimization problem (3) due to the computational complexity and the lack of knowledge.

As discussed above, the main challenge in solving (3) comes from the difficulty of calculating a researcher's future payoffs, due to the coupling across time and researchers. To choose the optimal effort level at each time t, each researcher i holds a conjecture that its future value $\mathbb{E}\left\{(1-\delta_i)\sum_{\tau=t}^{\infty}\delta_i^{\tau-t}u_i(m,\theta_i^\tau,d^\tau,e_i^\tau,\boldsymbol{e}_{-i}^\tau)\right\}$ (i.e., its discounted average payoff after time t) is the following:

$$f_i(\alpha_i,\beta_i^t,\theta_i^t,d^t,e_i^t) \triangleq$$
$$\alpha_i \cdot \sum_{j\neq k_i}\left[m_{k_ij}(\pi(\theta_i^t,q_i(e_i^t)),d(\boldsymbol{\theta})_j^t)\cdot b_i(d(\boldsymbol{\theta})_j^t)\right] + \beta_i^t. \quad (4)$$

In the above conjecture function, the term

$$\sum_{j\neq k_i}\left[m_{k_ij}(\pi(\theta_i^t,q_i(e_i^t)),d(\boldsymbol{\theta})_j^t)\cdot b_i(d(\boldsymbol{\theta})_j^t)\right]$$
$$=\sum_{j\neq k_i}\left[m_{k_ij}(\theta_i^{t+1},d(\boldsymbol{\theta})_j^t)\cdot b_i(d(\boldsymbol{\theta})_j^t)\right] \quad (5)$$

is the expected benefit that reviewer i will get in time $t+1$, assuming that the others' ratings remain the same (i.e., $\theta_j^{t+1}=\theta_j^t$ for all $j\neq i$).

Each researcher i holds such a conjecture for two reasons. First, it cannot predict the others' effort levels or future ratings. Hence, it holds a conjecture that *the others' ratings remain the same*, or equivalently, that the others' ratings precisely reflect their review quality, namely $d(\boldsymbol{\theta})_{k_j}^t = q_j(e_j^t)$. Second, it conjectures that its future value is an *affine function of its expected benefit*. Both of the above conjectures are required to be true in the equilibrium to be defined later.

The coefficient α_i reflects how "optimistic" a researcher is about the rating mechanism. A researcher with a larger α_i "believes in" the rating mechanism more, because it anticipates a higher future value given the expected benefit. The coefficient β_i^t is updated in each time slot by researcher i, such that the conjectured future value converges to the true future value in the equilibrium.

Then at each time t, each researcher i simply solves the following static problem for its optimal effort level $e_i^{t,*}$:

$$e_i^{t,*} = \arg\max_{e_i\in\mathbb{R}_+}\quad (1-\delta_i)\cdot u_i(m,\theta_i^t,d^t,e_i,\boldsymbol{e}_{-i})$$
$$+\quad \delta_i\cdot f_i(\alpha_i,\beta_i^t,\theta_i^t,d^t,e_i). \quad (6)$$

Each researcher i has all the information needed to solve the above static optimization problem.

Definition 1 (Conjectural Equilibrium [5]): Given any matching rule m and any rating update rule π, a conjectural equilibrium (CE) is a triple $\{\theta_i^*,e_i^*,\beta_i^*\}_{i\in\mathcal{N}}$ that satisfies:

- Incentive compatibility constraints: for all $i\in\mathcal{N}$,

$$e_i^* = \arg\max_{e_i\in[0,e_i^{\max}]}\quad (1-\delta_i)\cdot u_i(m,\theta_i^*,d^*,e_i,\boldsymbol{e}_{-i}^*)$$
$$+\quad \delta_i\cdot f_i(\alpha_i,\beta_i^*,\theta_i^*,d^*,e_i), \quad (7)$$

- Stable and correct ratings: for all $i\in\mathcal{N}$, $\theta_i^* = q_i(e_i^*)$,
- Consistent conjectures: for all $i\in\mathcal{N}$,

$$f_i(\alpha_i,\beta_i^*,\theta_i^*,d^*,e_i^*) = u_i(m,\theta_i^*,d^*,\boldsymbol{e}^*). \quad (8)$$

In the definition, the incentive compatibility (IC) constraints ensure that the effort level e_i^* is the best response of each researcher i, such that researcher i is in its self-interest to choose e_i^*. A CE also requires that each researcher's rating truly reflects its review quality at the equilibrium effort level e_i^*, and hence each researcher's rating is stable, namely $\pi(\theta_i^*,q_i(e_i^*))=\theta_i^*$. Finally, a CE requires that each researcher's conjecture about its future value is correct.

There may be multiple CEs. As a designer, it is desirable that the system will converge to a CE from any initial rating profile. The convergence is important, because the designer can distinguish the true review quality of the reviewers at the equilibrium. The choice of the matching rule plays an important role in ensuring the convergence to a CE.

D. The Design Problem Formulation

The designer's problem is to maximize the equilibrium review quality. We write the designer's objective as a function of the equilibrium review quality $W(q_1(e_1^*),\ldots,q_N(e_N^*))$. Then the designer problem can be defined as

$$\max_{m,\pi}\quad W(q_1(e_1^*),\ldots,q_N(e_N^*)) \quad (9)$$
$$s.t.\quad \{\theta_i^*,e_i^*,\beta_i^*\}_{i\in\mathcal{N}} \text{ is a CE under } m,\pi.$$

Note that the designer does not maximize the social welfare, namely the total benefit minus cost of the researchers. This is because the designer here is the editorial manager of a journal or a research community. It is more natural from an editorial manager's perspective to maximize the total review quality. The editorial manager may not care about the cost of reviewing paper; in fact, it would actually like to elicit more effort from the reviewers, resulting in higher costs.

IV. CONVERGENCE TO CONJECTURAL EQUILIBRIA

In this section, we consider general matching rules, and provide important guidelines for designing the matching rules. As discussed before, we would like to have a matching rule under which the system will converge to a CE from any initial rating profile under best response dynamics. In this way, the designer can distinguish the true review quality of the reviewers at the equilibrium. Before discussing the properties of the matching rules that ensure the convergence, we first describe the best response dynamics.

At each time slot t, the best response dynamics consist of the following three updates:

$$e_i^t = \arg\max_{e_i\in[0,e_i^{\max}]}\quad (1-\delta_i)\cdot u_i\left(m,\theta_i^t,d^t,e_i,\boldsymbol{e}_{-i}^t\right)$$
$$+\delta\cdot f_i\left(\alpha_i,\beta_i^t,\theta_i^t,d^t,e_i\right); \quad (10)$$

$$\theta_i^{t+1} = \begin{cases} (1-\mu) \cdot \theta_i^t + \mu \cdot q_i\left(e_i^t\right) & \text{if } i \text{ reviewed} \\ \theta_i^t & \text{otherwise} \end{cases} \quad ; \quad (11)$$

$$\begin{aligned} \beta_i^{t+1} &= u_i\left(m, \theta_i^t, d(\boldsymbol{\theta})^t, \boldsymbol{e}^t\right) \\ &\quad - \alpha_i \cdot \sum_{j \neq i} \left[m_{k_i j}\left(\theta_i^{t+1}, d(\boldsymbol{\theta})_j^t\right) \cdot b_i\left(d(\boldsymbol{\theta})_j^t\right) \right]. \end{aligned} \quad (12)$$

The update of effort levels in (10) and the update of ratings in (11) are the same as (6) and (1), respectively. They are rewritten here for the convenience of reference. When determining the effort level in (10), although the current payoff $u_i\left(m, \theta_i^t, d^t, e_i, \boldsymbol{e}_{-i}^t\right)$ depends on the others' effort levels $\boldsymbol{e}_{-i}^t$, the current payoff can be separated into the benefit which depends only on the others' effort $\boldsymbol{e}_{-i}^t$, and the cost which depends only on researcher i's own effort e_i. Hence, when solving (10), researcher i can treat the benefit as a constant, and consider only the cost, which depends on its own effort level and is known to researcher i.

The update of the parameter β_i in (12) ensures that the conjectured future payoff equals to the current payoff, namely $f_i\left(\alpha_i, \beta_i^{t+1}, \theta_i^t, d^t, e_i^t\right) = u_i\left(m, \theta_i^t, d(\boldsymbol{\theta})^t, \boldsymbol{e}^t\right)$. When the system converges to a CE $\{\theta_i^*, e_i^*, \beta_i^*\}_{i \in \mathcal{N}}$, we will have $f_i\left(\alpha_i, \beta_i^*, \theta_i^*, d^*, e_i^*\right) = u_i\left(m, \theta_i^*, d(\boldsymbol{\theta})^*, \boldsymbol{e}^*\right)$, which ensures that the third requirement of "consistent conjectures" in the definition of CE is fulfilled.

Next, we will provide the design guidelines on the matching rules, such that the above dynamics (10)–(12) always converge to a CE from any initial ratings. In fact, the design guideline is simple and intuitive: the matching rule should ensure that each researcher's expected benefit is concave and increasing in its own rating.

Definition 2 (Desirable Matching Rules): We say that a matching rule m is *desirable*, if under any rating profile $\boldsymbol{\theta}$,

- each researcher i's (conjectured) expected benefit from the review of its paper, namely

$$\sum_{j \neq k_i} \left[m_{k_i j}(d(\boldsymbol{\theta})_{k_i}, d(\boldsymbol{\theta})_j) \cdot b_i(d(\boldsymbol{\theta})_j) \right],$$

is concave and increasing in its own rating θ_i;
- each researcher i's expected number of papers to review is positive and fixed, namely

$$\sum_{j \neq k_i} m_{k_i j}(d(\boldsymbol{\theta})_j, d(\boldsymbol{\theta})_{k_i}) = M > 0.$$

The requirements of concavity and monotonicity are very reasonable. The expected benefit should be increasing in one's rating, such that one has incentives to exert high effort levels to increase its rating. In addition, if the expected benefit is concave in one's rating, since the marginal benefit is decreasing, one will not dramatically increase its effort level, which facilitates the convergence. The requirement of a fixed number of papers to review ensures the fairness among the reviewers across time.

Despite the simplicity of the requirements for desirable matching rules, we are able to prove its convergence under proper rating update rules.

Theorem 1 (Convergence): Under any desirable matching rule, starting from any initial $\boldsymbol{\theta}^0$, there exists $\bar{\mu} > 0$ such that under any small step size $\mu \in (0, \bar{\mu}]$ in the rating update rule, the system will converge to a CE through updates (10)–(12).

Proof: See [8]. ∎

V. DESIGN OF MATCHING RULES

The matching rule is the critical component of our design. In this section, we will first prove that some naive matching rules, which are reminiscent of how the papers are assigned in some of the current review systems, are inefficient. Next we propose a baseline matching rule, and analyze the properties of this baseline rule in detail. Finally, we extend this baseline rule in two directions.

A. Inefficiency of Naive Matching Rules

We show the inefficiency of the matching rules that do not consider the ratings of the researchers when assigning their papers. To some extent, such matching rules are reminiscent of how the papers are assigned in some existing systems. In such systems, a paper is assigned either to a reviewer randomly, or to a reviewer based only on the reviewer's rating (e.g., its past history of reviewing). As we will show next, such matching rules are inefficient in the sense that they cannot incentivize reviewers to exert high effort.

Proposition 1: Under any matching rule that is independent of the researcher's rating, namely $m_{kj}(d(\boldsymbol{\theta})_k, d(\boldsymbol{\theta})_j) = m_{k'j}(d(\boldsymbol{\theta})_{k'}, d(\boldsymbol{\theta})_j)$ for any k, k', and $\boldsymbol{\theta}$, there is a unique CE, in which $e_i^* = 0$ and $\theta_i^* = q_i(e_i^*) = 0$ for all i.

Proof: See [8]. ∎

The above proposition underlines the importance of designing matching rules that take into account not only the ratings of the reviewers, but also those of the researchers.

B. Design of The Baseline Matching Rule

The baseline matching rule works as follows:

1) For the researchers with the same rating, match their papers among themselves using any one-to-one mapping that does not match one's paper to itself.
2) For any researcher i with a distinct rating (i.e., no other researcher has the same rating),
   a) If it has the highest rating (i.e., $k_i = 1$), match its paper to a reviewer with the second highest rating.
   b) If it has the lowest rating (i.e., $k_i = N$), match its paper to a reviewer with the second lowest rating with probability $\frac{d(\boldsymbol{\theta})_N}{d(\boldsymbol{\theta})_{N-1}}$. Hence, its paper gets no reviewer with probability $1 - \frac{d(\boldsymbol{\theta})_N}{d(\boldsymbol{\theta})_{N-1}}$.
   c) If $1 < k_i < N$, match its paper to its two "neighbors" with the following probabilities (which sum up to 1):

$$m_{k_i, k_i-1}(d(\boldsymbol{\theta})_{k_i}, d(\boldsymbol{\theta})_{k_i-1}) = \frac{d(\boldsymbol{\theta})_{k_i} - d(\boldsymbol{\theta})_{k_i+1}}{d(\boldsymbol{\theta})_{k_i-1} - d(\boldsymbol{\theta})_{k_i+1}},$$

and

$$m_{k_i, k_i+1}(d(\boldsymbol{\theta})_{k_i}, d(\boldsymbol{\theta})_{k_i+1}) = \frac{d(\boldsymbol{\theta})_{k_i-1} - d(\boldsymbol{\theta})_{k_i}}{d(\boldsymbol{\theta})_{k_i-1} - d(\boldsymbol{\theta})_{k_i+1}}.$$
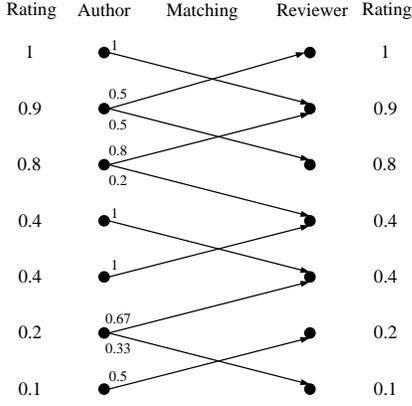
Fig. 1. An illustration of the baseline matching rule. There are 7 researchers. The paper of the researcher with the highest rating is matched to the reviewer with the second highest rating. The paper of the researcher with the lowest rating is matched to the reviewer with the second lowest rating with probability 0.5. The two researchers with the same rating 0.4 are matched to each other. The rest are matched to their two nearest neighbors with probabilities inversely proportional to the distances in ratings.

The above matching rule matches the researchers with the same rating to each other. For a researcher with a distinct rating, it matches its paper with its two nearest "neighbors" with probabilities that depend on how close its rating is to its neighbors' ratings.

1) The Choice of Initial Ratings: In the considered system, it is important to choose the initial ratings correctly, because under different initial ratings, the system may converge to different CEs. Since the designer has no knowledge about the researchers at the beginning, it is reasonable to assign the same initial rating to all the researchers for fairness. In this case, the following proposition tells us that we should not assign the initial rating too low.

Proposition 2: There always exists a rating $\underline{\theta}$, such that any initial rating profile with the same rating $\theta^0 \leq \underline{\theta}$ for all the researchers is the equilibrium rating profile, and that each researcher i chooses an equilibrium effort level e_i^* such that $q_i(e_i^*) = \theta^0$.

Proof: See [8]. ∎

Proposition 2 indicates that we should choose a high enough initial rating. The key reason is that no researcher has incentive to reach a higher rating than the initial one, because it will get a *distinct* highest rating in this case. Then its benefit will stay the same, while the cost will be higher, compared to choosing an effort level such that its rating remains the same as the initial rating. In other words, the initial rating determines the highest review quality produced by each researcher. We show that when the initial rating is low enough, it is the best response to choose an effort level e_i^* that satisfies $q_i(e_i^*) = \theta^0$.

2) Convergence: It is useful to classify researchers into types based on their cost, review quality, and benefit functions, etc. We define researchers of a certain type as follows.

Definition 3 (Types): The researchers of the same *type* have the same normalized marginal benefit to cost ratio, defined as $\frac{\delta_i \alpha_i q_i'(\cdot)}{(1-\delta_i)c_i'(\cdot)}$, the same review quality function $q_i(\cdot)$,

and the same marginal benefit function $b_i'(\cdot)$.

Definition 4 (Ordering of Capability): A researcher i is *more capable* than a researcher j, if

$$\frac{\delta_i \alpha_i q_i'(e)}{(1-\delta_i)c_i'(e)} \geq \frac{\delta_j \alpha_j q_j'(e)}{(1-\delta_j)c_j'(e)}, \forall e,$$
$$q_i(e) > q_j(e), \forall e,$$
$$b_i'(\theta) \geq b_j'(\theta), \forall \theta.$$

Definition 3 identifies "types" of researchers, in the sense that the researchers of the same type will always choose the same effort level and hence get the same rating. Definition 4 gives an ordering of the researchers in terms of their "capability". We will prove that a more capable researcher indeed produces higher-quality review and gets higher ratings.

In the rest of this section, we make the following assumption about the population size of the community.

Assumption 2 (Large Population): There is more than one researcher of each type.

Assumption 2 is reasonable in practice, since the number of researchers is indeed large. Given the same initial rating, the researchers of the same type will choose the same best response effort level, and hence have the same rating. Assumption 2 ensures that for each researcher, there is always another researcher with the same rating. According to Clause 1) in the baseline matching rule, each researcher will always have exactly one paper to review all the time.

Theorem 2: Suppose that the large population assumption (Assumption 2) holds. Then we have

- the baseline matching rule is a desirable matching rule;
- starting from any initial rating profile, there exists $\bar{\mu} > 0$ such that under any small step size $\mu \in (0, \bar{\mu}]$ in the rating update rule, the system will converge to a CE through the best response dynamics (10)–(12);
- if the researchers have the same initial rating, at any point in the best response dynamics (10)–(12), more capable researchers will always find it in their self-interest to produce higher review quality, and thus have higher ratings than less capable researchers.

Proof: See [8]. ∎

Theorem 2 ensures the convergence of the best response dynamics to a CE. In fact, we can say something stronger about the best response dynamics: a more capable researcher has a higher rating than a less capable researcher *at any point* in the best response dynamics. This means that the rating mechanism can successfully distinguish the researchers of different types, and rank them in the correct order. Note that more capable researchers produce higher-quality review in their self-interest, as a result of maximizing their own payoffs; they are not obliged to do so by the designer.

C. Two Classes of Extended Matching Rules

Previously, we focused on the baseline matching rule. The baseline matching rule is able to incentivize the researchers to exert high effort levels, by increasing the benefit obtained by a researcher when its rating increases. Now we extend the baseline rule in two different ways, both of which result
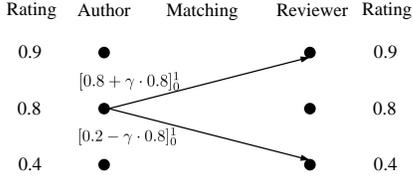
Fig. 2. An illustration of the first extension of matching rules. We only show the matching of the researcher with rating 0.8, who is matched to its two nearest neighbors with different probabilities than in the baseline rule.

Fig. 3. An illustration of the second extension of matching rules. We only show the matching of the researcher with rating 0.8, who is matched to possibly its four neighbors, instead of two nearest neighbors in the baseline rule.

in a class of matching rules that allow us to tune the reward and/or punishment provided by the matching rules.

In the first extension, we assign the probability that the paper of a researcher with a distinct rating is assigned to the neighbors. In particular, the matching rule is parametrized by γ such that any researcher i with a distinct rating and with $k_i \in [2, N-1]$ is matched to its neighbors with the following probabilities:

$$m_{k_i, k_i-1}(d(\boldsymbol{\theta})_{k_i}, d(\boldsymbol{\theta})_{k_i-1}) = \left[\frac{d(\boldsymbol{\theta})_{k_i} - d(\boldsymbol{\theta})_{k_i+1}}{d(\boldsymbol{\theta})_{k_i-1} - d(\boldsymbol{\theta})_{k_i+1}} + \gamma \cdot \theta_i \right]_0^1,$$

and

$$m_{k_i, k_i+1}(d(\boldsymbol{\theta})_{k_i}, d(\boldsymbol{\theta})_{k_i+1}) = \left[\frac{d(\boldsymbol{\theta})_{k_i-1} - d(\boldsymbol{\theta})_{k_i}}{d(\boldsymbol{\theta})_{k_i-1} - d(\boldsymbol{\theta})_{k_i+1}} - \gamma \cdot \theta_i \right]_0^1,$$

where $[\cdot]_0^1 \triangleq \min\{\max\{\cdot, 0\}, 1\}$.

We illustrate this extension in Fig. 2.

We can see that when $\gamma > 0$ ($\gamma < 0$), the resulting matching rule rewards (punishes) the researcher by increasing its probability of being matched to the higher-rating (lower-rating) neighbor. When $\gamma = 0$, it reduces to the baseline matching rule.

In the second extension, we allow a researcher to be matched to a reviewer with even higher or even lower ratings than its nearest neighbors. In particular, the matching rule is parametrized by $\gamma_r \in [0,1]$ and $\gamma_p \in [0,1]$. Then any researcher i with a distinct rating and with $k_i \in [3, N-2]$ is matched to its neighbors and neighbors of neighbors with the following probabilities:

$$m_{k_i, k_i-1}(d(\boldsymbol{\theta})_{k_i}, d(\boldsymbol{\theta})_{k_i-1}) = \frac{d(\boldsymbol{\theta})_{k_i} - d(\boldsymbol{\theta})_{k_i+1}}{d(\boldsymbol{\theta})_{k_i-1} - d(\boldsymbol{\theta})_{k_i+1}} \cdot (1 - \gamma_r),$$

$$m_{k_i, k_i-2}(d(\boldsymbol{\theta})_{k_i}, d(\boldsymbol{\theta})_{k_i-2}) = \frac{d(\boldsymbol{\theta})_{k_i} - d(\boldsymbol{\theta})_{k_i+1}}{d(\boldsymbol{\theta})_{k_i-1} - d(\boldsymbol{\theta})_{k_i+1}} \cdot \gamma_r,$$

and

$$m_{k_i, k_i+1}(d(\boldsymbol{\theta})_{k_i}, d(\boldsymbol{\theta})_{k_i+1}) = \frac{d(\boldsymbol{\theta})_{k_i-1} - d(\boldsymbol{\theta})_{k_i}}{d(\boldsymbol{\theta})_{k_i-1} - d(\boldsymbol{\theta})_{k_i+1}} \cdot (1 - \gamma_p),$$

$$m_{k_i, k_i+2}(d(\boldsymbol{\theta})_{k_i}, d(\boldsymbol{\theta})_{k_i+2}) = \frac{d(\boldsymbol{\theta})_{k_i-1} - d(\boldsymbol{\theta})_{k_i}}{d(\boldsymbol{\theta})_{k_i-1} - d(\boldsymbol{\theta})_{k_i+1}} \cdot \gamma_p.$$

We illustrate this extension in Fig. 3.

We can see that the parameters γ_r and γ_p reflect to what extent the researchers are rewarded and punished, respectively. When $\gamma_r = \gamma_p = 0$, the matching rule reduces to the baseline rule. When $\gamma_r = 1$ ($\gamma_p = 1$), the researcher is rewarded (punished) by being matched to a reviewer with the next higher (lower) rating.

It is interesting to ask under each class of extended matching rules, which matching rule is optimal in terms of
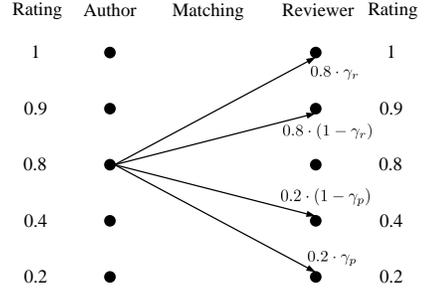
the equilibrium review quality? We first define the notion that one matching rule is "better" than the other.

Definition 5: We say that a matching rule m' is "better" than another matching rule m, if for any equilibrium rating profile $\boldsymbol{\theta}^*$ under m, we can find an equilibrium rating profile $\boldsymbol{\theta}^{*\prime}$ under m' that satisfies $\boldsymbol{\theta}^{*\prime} > \boldsymbol{\theta}^*$.

The following theorem tells us how to design an extended matching rule that is better than the baseline rule.

Theorem 3: Suppose that the large population assumption (Assumption 2) holds. Then we have:

- in the first extension, there exists a $\gamma > 0$ (i.e., rewarding) under which the extended rule is better than the baseline rule;
- in the second extension, there exists $\gamma_r = 0$ and $\gamma_p > 0$ (i.e., punishing) under which the extended rule is better than the baseline rule.

Proof: See [8]. ∎

Theorem 3 tells us if we reward or punish by assigning higher or lower probabilities of being matched to the higher-rating neighbor, it is beneficial to reward. On the contrary, if we reward or punish by creating the possibility of being assigned to the next higher- or lower-rating neighbors, it is beneficial to punish. Note that we can get the benefit only when we set the correct parameters in the extended matching rules. The technical reason is that we want to increase the marginal expected benefit when a researcher's rating is changed, in order to give more incentive for them to exert high effort levels. The message delivered by our result is that, we should carefully design the reward and punishment mechanism based on the way we reward and punish.

VI. SIMULATION RESULTS

We consider a system with 5 types of researchers. There are 100 researchers of each type. All the researchers have the same patience $\delta_i = 0.8$, the same cost function $c_i(e_i) = e_i^2$, and the same benefit function $b_i(\theta) = -\theta^2 + 2\theta$. Different types of researchers have different quality function $q_i(e_i) = p_i \cdot e_i$, where $p_i = 0.2, 0.4, 0.6, 0.8,$. They also have different optimism $\alpha_i = 1.1, 1.2, 1.3, 1.4, 1.5$.

1) Convergence and Performance Improvement: Fig. 4 shows the convergence to the conjectured equilibrium from the initial rating 1.0. As predicted by Theorem 2, the
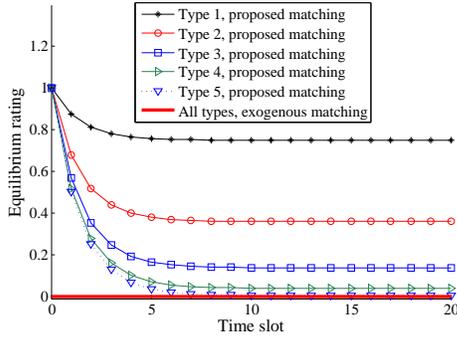
Fig. 4. Convergence to the conjectural equilibrium.
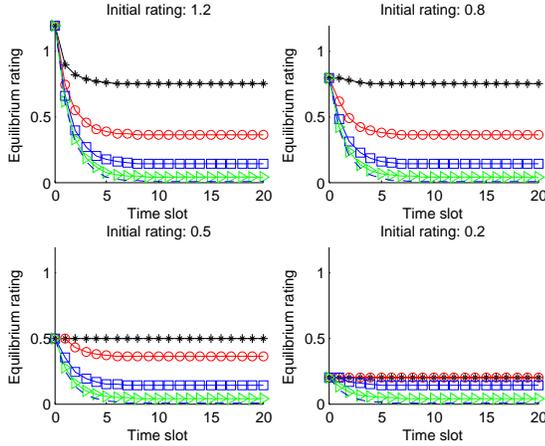
Fig. 5. Equilibria under different initial ratings.

equilibrium rating and review quality is ordered according to the types of the researchers. Fig. 4 also shows that under exogenous matching rules that do not depend on researchers' ratings, the reviewers exert lowest efforts all the time. Hence, the proposed endogenous matching greatly improves the performance of the review system.

2) Impact of Initial Ratings: In Fig. 5, we examine the impact of the initial rating. We can see that when the initial rating is high enough (1.2 and 0.8), the system converges to the same equilibrium. When the initial rating is low (0.5), some types may stay at the initial rating. This is consistent with Proposition 1, which indicates that when the initial rating is sufficiently small, the initial rating is the equilibrium rating for all the researchers.

3) Different Matching Rules: We compare the sum review quality and the social welfare (i.e., the total benefit minus cost) at the equilibrium under different matching rules.

In Table I, we evaluate the first extension of matching rules

TABLE I

EQUILIBRIUM REVIEW QUALITY AND SOCIAL WELFARE UNDER THE
FIRST EXTENSION OF MATCHING RULES.

γ	-0.2	-0.1	-0.05	0	0.05	0.1	0.2
Sum review quality	0.64	0.91	0.96	1.29	1.28	**1.36**	1.28
Social welfare	1.37	1.58	**1.59**	1.44	1.45	1.46	1.55

TABLE II

EQUILIBRIUM REVIEW QUALITY AND SOCIAL WELFARE UNDER THE
SECOND EXTENSION OF MATCHING RULES.

(γ_r, γ_p)	(0, 0)	(0, .5)	(0, 1)	(.5, 0)	(.5, .5)	(.5, 1)
Sum review quality	1.29	1.31	**1.40**	1.11	1.28	1.33
Social welfare	1.44	1.41	1.35	1.27	**1.57**	1.43

under different parameters γ. We can see that in our setting, the optimal γ should be 0.1, which results in the highest sum review quality. This is consistent with our theoretical results: we can find a rewarding matching rule that outperforms the baseline rule. It is worth mentioning that the matching rule that maximizes the sum review quality may not be the one that maximizes the social welfare. This is reasonable, because higher review quality also results in higher cost.

In Table II, we evaluate the second extension of matching rules under different parameters γ_r and γ_p. We can see that the optimal sum review quality is achieved when $\gamma_r = 0$ and $\gamma_p = 1$, which is a matching rule that punishes to the most severe extent. The threat of being matched to an even lower-rating reviewer provides more incentive for researchers to exert high effort. Again, such a matching rule does not result in the optimal social welfare. The optimal social welfare is achieved when $\gamma_r = 0.5$ and $\gamma_p = 0.5$, where the researchers are also rewarded for good behaviors.

VII. CONCLUSION

We studied the problem of effort elicitation in peer review systems. We modeled the two key features in such systems: namely moral hazard (i.e., the unobservable effort by the reviewers) and adverse selection (i.e., the unidentifiable quality of the reviewers). We proposed a rating and matching mechanism to identify the reviewers of different types, and elicit the appropriate amount of effort from different reviewers. We extensively studied the design of matching rules, in terms of the initial ratings, the convergence, and the equilibrium ratings and review effort. We also studied the extensions to different classes of matching rules, and proved the efficiency of different reward and punishment mechanisms under different matching rules.

REFERENCES

[1] D. Kurokawa, O. Lev, J. Morgenstern, and A. D. Procaccia, "Impartial Peer Review," *Technical Report*, 2014. Available: http://www.cs.cmu.edu/ arielpro/papers/peer_review.pdf

[2] M. Kandori, "Social norms and community enforcement," *Review of Economic Studies*, vol. 59, no. 1, pp. 63 – 80, 1992.

[3] Y. Zhang, J. Park, and M. van der Schaar, "Rating protocols for online communities," *ACM Transactions on Economics and Computation*, vol. 2, no. 1, 2013.

[4] Y. Xiao and M. van der Schaar, "Socially-optimal design of service exchange platforms with imperfect monitoring," *Technical Report*, 2013. Available: http://arxiv.org/abs/1310.2323

[5] F. H. Hahn, "Exercises in conjectural equilibria," *The Scandinavian Journal of Economics*, Vol. 79, No. 2, pp. 210-226, 1977.

[6] A. Abdulkadiroğlu and T. Sönmez, "School choice: A mechanism design approach," *American Economic Review*, vol. 93, no. 3, 2003.

[7] A. E. Roth, T. Sönmez, and M. Utku Üver, "Pairwise kidney exchange," *Journal of Economic Theory*, vol. 125, no. 2, 2005.

[8] Yuanzhang Xiao, Florian Dörfler, and Mihaela van der Schaar, "Rating and matching in peer review systems," *Technical Report*. Available: http://medianetlab.ee.ucla.edu/papers/PeerReview.pdf