Incentive Design in Peer Review: Rating and Repeated Endogenous Matching

Yuanzhang Xiao, Member, IEEE, Florian Dörfler, Member, IEEE, and Mihaela van der Schaar, Fellow, IEEE

Abstract—Peer review (e.g., grading assignments in Massive Open Online Courses (MOOCs), academic paper review) is an effective and scalable method to evaluate the products (e.g., assignments, papers) of a large number of agents when the number of dedicated reviewing experts (e.g., teaching assistants, editors) is limited. Peer review poses two key challenges: 1) identifying the reviewers’ intrinsic capabilities (i.e., adverse selection) and 2) incentivizing the reviewers to exert high effort (i.e., moral hazard). Some works in mechanism design address pure adverse selection using one-shot matching rules, and pure moral hazard was addressed in repeated games with exogenously given and fixed matching rules. However, in peer review systems exhibiting both adverse selection and moral hazard, one-shot or exogenous matching rules do not link agents’ current behavior with future matches and future payoffs, and as we prove, will induce myopic behavior (i.e., exerting the lowest effort) resulting in the lowest review quality. In this paper, we propose for the first time a solution that simultaneously solves adverse selection and moral hazard. Our solution exploits the repeated interactions of agents, utilizes ratings to summarize agents’ past review quality, and designs matching rules that endogenously depend on agents’ ratings. Our proposed matching rules are easy to implement and require no knowledge about agents’ private information (e.g., their benefit and cost functions). Yet, they are effective in guiding the system to an equilibrium where the agents are incentivized to exert high effort and receive ratings that precisely reflect their review quality. Using several illustrative examples, we quantify the significant performance gains obtained by our proposed mechanism as compared to existing one-shot or exogenous matching rules.

Index Terms—Peer review, rating mechanisms, matching mechanisms, incentive design

1 INTRODUCTION

Peer review serves as an effective and scalable method for performance evaluation in systems where the products to evaluate significantly outnumber the dedicated reviewing experts. One example of such systems is Massive Open Online Courses (MOOCs), where the number of students enrolled in a course is in the order of tens of thousands and by far exceeds the number of teaching assistants [2], [3], [4], [5]. Another example is academic paper review, where the number of papers submitted to a journal by far exceeds the number of (associate) editors. Since the proposed mechanism can be applied to general peer review settings, we keep the discussion in this paper general.

Peer review systems pose two key challenges. First, the reviewers have different intrinsic capabilities (e.g., their review quality functions, benefit and cost functions), which are unknown. Hence, one challenge is how to identify their unknown intrinsic capabilities; this is known in the game theory literature as the adverse selection problem. Second, the reviewers can choose to exert different levels of (costly) effort (e.g., time and energy spent in reviewing), which is unobservable. Hence, the other challenge is how to incentivize reviewers to exert high effort; this is known in the game theory literature as the moral hazard problem. A reviewer’s ultimate review quality is determined by her intrinsic capabilities and effort. If the capabilities are unknown but the effort is observable (i.e., pure adverse selection), there is hope to identify their capabilities through mechanism design. If the effort is unobservable but the capabilities are known (i.e., pure adverse selection), there is hope to incentivize high effort through social norms. However, in the presence of both adverse selection and moral hazard, the problem becomes significantly more challenging. In fact, no existing work has addressed this problem systematically.

A natural candidate for solving the pure adverse selection problem is to use matching mechanisms [6], [7], [8]. Matching mechanisms aim to efficiently allocate resources (e.g., hospitals, or reviewers in our setting) to agents (e.g., medical students) or their products (e.g., assignments or papers in our setting). Existing matching mechanisms assume that the quality of resources depend only on the types of the agents who provide and receive the resource, but not on the providers’
effort. In other words, there is no moral hazard problem. As a result, they focus on one-shot interactions and design one-shot matching rules (i.e., each agent is matched only once). However, their assumption does not hold in peer review systems, where the review quality depends crucially on the reviewers’ effort. We prove that under one-shot matching rules, agents will behave myopically by choosing the lowest effort (i.e., free-riding), because their current effort does not affect future matches and their future payoffs. Hence, the system performance (in terms of the total review quality) is the worst since one-shot matching does not address the moral hazard problem.

One way to address the pure moral hazard problem is to use social norms [10], where a central agency assigns each agent with a rating that summarizes her past behavior and recommends a “norm” (i.e., desired behavior) that rewards agents with good ratings and punishes those with bad ratings. In this way, the agents are incentivized to conform with the social norm (e.g., exert high effort in our setting), even when they are randomly matched to each other based on some exogenous matching rule. However, existing works on social norms assume that the agents are homogenous. This assumption does not hold in peer review systems, because different reviewers have different intrinsic capabilities. Ideally, the central agency should recommend different norms to agents of different capabilities; in practice, it cannot do this since the capabilities are unknown. In summary, existing social norms [10], [11], [12], [13] do not deal with the adverse selection problem that is present in peer review.

This paper proposes the first mechanism to simultaneously solve the adverse selection and moral hazard problems in peer review. Our proposed mechanism exploits the repeated interaction among agents (i.e., by submitting multiple assignments or papers over time), and assigns the agents with ratings, which are summaries of their past review quality. Unlike the works on social norms [10], [11], [12], [13], we do not recommend desired behavior (i.e., a social norm) to the agents, because they have unknown, different capabilities and thus computing a recommended social norm is impossible. Instead, we propose rules for repeated matching that endogenously depend on agents’ ratings. Unlike existing one-shot [6], [7], [8] or exogenous [10], [11], [12], [13] matching rules, our proposed repeated endogenous matching rules provide strong incentives for agents to exert high effort, because the agents’ behaviors affect their future ratings and hence, their future matches and future payoffs.

We provide design guidelines for endogenous matching rules that are easy to implement without knowledge of agents’ private information (e.g., their benefit, review quality, and cost functions), yet powerful enough to guide the system to desirable equilibria. In particular, in the equilibrium the agents find it in their self-interest to exert high effort, and receive ratings that truly reflect their capabilities. We also provide case studies on specific matching rules with different reward/punishment schemes. We show that different reward/punishment schemes lead to different optimal matching rules, which stresses the importance of tailoring matching rules to reward/punishment schemes. Simulation results demonstrate large performance improvement over existing matching rules.

In the following, we discuss related works in Section 2. Then we describe the model and formulate the design problem in Section 3. We study general matching rules in Section 4, and the baseline matching rule and its extensions in Section 5. Section 6 demonstrates the efficiency of our proposed mechanisms. Finally, Section 7 concludes the paper.

2 Related Works

2.1 Pure Adverse Selection

The pure adverse selection problem is the focus of a huge literature on matching in resource allocation (e.g., allocation of schools to applicants [6], [7]) and exchange (e.g., kidney exchange [8]). These works ignore the moral hazard problem. In particular, they do not consider “effort”. Once an agent (e.g., an applicant) is matched to another (e.g., a school), the benefit (obtained by this applicant) is fixed. In contrast, in our work, the review quality depends crucially on the reviewer’s effort. This additional moral hazard problem, when ignored, will significantly degrade the system performance (in terms of the total review quality).

Since these works [6], [7], [8] ignore moral hazard, their matching rules are one-shot (i.e., match each agent only once). In contrast, our matching rules are repeated and changing over time based on agents’ ratings. Hence, our matching rules can incentivize agents to exert high effort to obtain better ratings and thus, favorable future matches.

2.2 Pure Moral Hazard

The pure moral hazard problem has been studied in repeated game theory, where anonymous agents are randomly matched to interact with each other [10], [11], [12], [13], as in our work. However, these works [10], [11], [12], [13] focus on the pure moral hazard problem, and ignore the adverse selection problem by assuming homogenous agents. In this work, we assume heterogeneous agents and deal with both the moral hazard and adverse selection problems. In [10], [11], [12], [13], due to the homogeneity of agents, binary ratings are usually sufficient to identify whether an agent has behaved well or badly. In contrast, in this work the rating is continuous, such that the rating mechanism can identify not only whether an agent has behaved well or badly, but also its review quality.

Another key difference is that we design matching rules that endogenously depend on agents’ ratings and directly affect their incentives, while the matching rules in [10], [11], [12], [13] are fixed and exogenously given. In our setting, we will prove that the latter type of matching rules will result in the lowest review quality in the equilibrium.

2.3 Other Works on Peer Review

There are other works on peer review systems but with different problems to solve. In [16] an auction market for paper submission is proposed that precedes the actual peer review. There are works focusing on how to aggregate reviewers’ scores/ratings to obtain a final score/rating that accurately reflects the true quality of the assignments (in MOOCs [3], [4], [5], the proposals (in NSF proposal reviewing [14]), or the papers (in academic peer review [15]). In contrast, our focus is to incentivize reviewers to exert high effort levels.

---

2. There are works on dynamic matching (see representative work [9]). However, matching is called dynamic due to the dynamic arrival and departure of the agents. Each agent is still matched only once.
3 Model

3.1 Basic Setup
Consider a peer review system with a set \( \mathcal{N} = \{1, \ldots, N\} \) of \( N \) agents. Each agent has its products reviewed by the other agents. An agent benefits from the review by its reviewer, and exerts effort in reviewing others’ products. A designer (e.g., the instructor in MOOCs) aims to design a mechanism that incentivizes the reviewers to produce high-quality reviews. The mechanism includes two parts: (i) the rating mechanism that assigns and updates a rating \( \theta_i \in \mathbb{R}_+ \) for each agent \( i \), and (ii) the matching rule that matches agents with reviewers (possibly based on the ratings). In the following, we write the rating profile, namely the ratings of every agent, as \( \theta = (\theta_1, \ldots, \theta_N) \). The rating profile is known only to the designer. We define the rating distribution, denoted by a vector \( d(\theta) \), as the ordered (from high to low) list of all the ratings. The rating distribution \( d(\theta) \) does not count multiple agents with the same rating. For example, if the rating profile is \( \theta = (3, 5, 5, 3) \), the rating distribution will be \( d(\theta) = (5, 3) \). Write \( K \) as the number of distinct ratings in \( \theta \) (i.e., the dimension of the vector \( d(\theta) \)), and \( k_i \) as agent \( i \)'s ranking (i.e., ordered position) in the rating distribution. In the above example, we have \( K = 2, k_1 = k_4 = 2, k_2 = k_3 = 1 \). Although \( K \) and \( k_i \) depend on \( \theta \), we write them simply as \( K \) and \( k_i \) for notational simplicity without causing confusion. Denote the \( k_i \)th element of the rating distribution by \( d(\theta)_{k_i} \). Then we have \( \theta_i = d(\theta)_{k_i} \). Finally, notice that the rating distribution does not disclose any information about the identities of the agents.

Note, importantly, that an agent’s rating indicates its review quality, not the quality of its product.

Time is slotted into \( t = 0, 1, 2, \ldots \). In each time slot \( t \), the entities in the system moves in the following order3:

- The designer publishes the rating distribution \( d(\theta) \), and informs agent \( i \) of its rating \( \theta_i \) and its ranking \( k_i \).
- Each agent submits its product to review.
- Each agent matches each agent \( i \)'s product to other agent(s) for review based on a probabilistic matching rule \( m_{k_i k_j} : (d(\theta)_{k_i}, d(\theta)_{k_j}) \rightarrow [0, 1] \). The matching rule determines the probability \( m_{k_i k_j} (d(\theta)_{k_i}, d(\theta)_{k_j}) \) that the agent with the \( k_i \)th highest rating is matched to the reviewer with the \( k_j \)th highest rating. From the definition we can see that the matching does not depend on agents’ identities.
- Each reviewer \( j \) exerts an effort level \( e_j^i \in [0, e_j^{\text{max}}] \), where \( e_j^{\text{max}} \) is \( j \)'s maximum effort level. Reviewer \( j \)'s review quality then depends on its effort as \( q_j(e_j^i) \), where \( q_j : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) is the review quality function. The designer is informed of each reviewer’s the review quality by the agents.
- Each agent \( i \) receives benefit \( b_i(q_j(e_j^i)) \) from reviewer \( j \)'s review, where \( b_i : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) is \( i \)'s benefit function, and incurs a cost of \( c_i(e_j^i) \) for reviewing a product, where \( c_i : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) is \( i \)'s cost function.
- The designer updates the agents’ ratings according to the rating update rule \( \pi : (\theta_j, q_j(e_j^i)) \rightarrow \theta_j^{t+1} \).

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 with a constant step size \( \mu \in (0, 1) \):

\[
\theta_j^{t+1} = \begin{cases} 
(1 - \mu) \cdot \theta_j^t + \mu \cdot q_j(e_j^t), & j \text{ has reviewed} \\
\theta_j^t, & \text{otherwise}.
\end{cases}
\]

(1)

We make the following remarks on the agents’ ratings. Each agent \( i \) has a maximum effort level \( e_i^{\text{max}} \) and hence has a maximum review quality \( q_i(e_i^{\text{max}}) \). Since the new rating is the convex combination of the old rating and the review quality, given any initial rating \( \theta_i^0 \), agent \( i \)'s rating can only be in the interval \( [0, \theta_i^{\text{max}}] \), where \( \theta_i^{\text{max}} \triangleq \max \{\theta_i^0, q_i(e_i^{\text{max}})\} \). In other words, the possible ratings of each agent \( i \) are contained in the compact set \( [0, \theta_i^{\text{max}}] \).

Throughout the paper, we make the following reasonable and standard assumptions on the monotonicity, convexity, concavity, and differentiability of our functions.

**Assumption 1 (Cost, Review Quality, and Benefit).** Each agent \( i \)'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 twice continuously differentiable in effort \( e_i \). In addition, \( c_i'(0) = 0 \).
- The review quality \( q_i(\cdot) \) is concave, strictly increasing, and twice continuously differentiable in effort \( e_i \). In addition, \( q_i'(0) \) exists and is bounded.
- The benefit \( b_i(\cdot) \) is strictly increasing, concave, and continuously differentiable in review quality \( q_j \). In addition, \( b_i'(0) \) exists and is bounded.
- We normalize \( c_i(0) = 0, q_i(0) = 0, \) and \( b_i(0) = 0 \).

The requirements on \( c_i(0), q_i(0), \) and \( b_i(0) \) are just normalizations and are imposed without loss of generality. The assumptions on convexity, monotonicity, and differentiability are technical, but make sense and are common in the literature. They are needed for our technical results.

3.2 Information – Who Knows What

3.2.1 The Designer
The designer receives reports of review quality \( q_j(e_j^t) \), and keeps the rating \( \theta_j^t \) for each agent \( i \) at each time slot \( t \). Hence, the designer knows the identity of the agent at the \( k_i \)th position of the rating distribution. However, it does not know the review quality functions \( q_j(\cdot) \), the benefit functions \( b_i(\cdot) \), or the cost functions \( c_i(\cdot) \).

3.2.2 Each Agent \( i \)
Each agent \( i \) 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 agents. It knows the matching rule \( m \) and the rating update rule \( \pi \). It also knows its own rating \( \theta_i^t \), the rating distribution \( d(\theta_i^t) \), and its position in the rating distribution \( k_i^t \), but does not know the others’ ratings or the identity of its reviewer.

---

3. Throughout the paper, the superscript \((\cdot)^t\) on a function refers to the derivative, and the superscript \((\cdot)\) refers to the variable under consideration at time point \( t \in \mathbb{Z}_+ \).
3.3 Payoffs and Equilibrium

In each time slot $t$, agent $i$’s expected payoff is its expected benefit from the reviewing of its product minus the expected cost of reviewing other agents’ products. We write agent $i$’s expected payoff as $u_i(m, \theta_i, d(\theta), e)$, which depends on the matching rule $m$, its own rating $\theta_i$, the rating distribution $d(\theta)$, and all the agents’ effort levels $e \triangleq (e_1, \ldots, e_N)$. The expected payoff can be calculated as the expected benefit minus the expected cost:

$$u_i(m, \theta_i, d(\theta), e) = \sum_{kj} m_{ik} \left( d(\theta)_{kj} \cdot b_i(q_j(e_j)) \right)$$

$$- \sum_{kj} m_{ik} \left( d(\theta)_{kj} \right) \cdot c_i(e_i).$$

(2)

Each agent $i$ aims to choose a sequence of effort levels over time to maximize the discounted average of expected payoffs, i.e., to solve the dynamic optimization problem below:

$$\max_{\{e'_t \in [0, \text{max}_t]\}^T} \mathbb{E} \left\{ (1 - \delta) \sum_{t=0}^{\infty} \delta^t u_i(m, \theta_i, d(\theta'), e'_t, e'_{t-1}) \right\},$$

(3)

where $e'_{t-1}$ is the effort levels chosen by all the agents other than $i$ at time $t$, and $\delta \in [0,1)$ is agent $i$’s discount factor. An agent’s discount factor reflects its patience. We take the expectation $\mathbb{E}\{\cdot\}$ because the rating update is random, namely an agent’s rating is either updated or kept the same depending on whether it has reviewed a product.

Note that the optimization problem (3) is very hard, if not impossible, to solve. The difficulty lies in the couplings of one agent’s decisions over time and with other agents’ decisions. First, the agent’s current 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 agent’s payoff is affected by the others’ decisions (through the benefit). However, since an agent has no knowledge about the others, it cannot predict the others’ decisions and the evolution of rating distributions. In summary, an agent cannot solve the optimization problem (3) due to computational complexity and lack of knowledge.

We propose a realistic behavioral model for the agents. To choose the optimal effort level at each time $t$, each agent $i$ holds a conjecture that its future value $\mathbb{E}\{(1 - \delta) \sum_{t=0}^{\infty} \delta^t u_i(m, \theta_i, d(\theta'), e'_t, e'_{t-1})\}$ (i.e., its discounted average payoff after time $t$) is the following:

$$f_i(\alpha_i, \beta'_i, d(\theta'), e'_t) = \tilde{b}_i(\theta'_i, d(\theta'), e'_t) + \beta'_t,$$

(4)

where $\tilde{b}_i(\theta'_i, d(\theta'), e'_t)$ is the conjectured expected benefit of agent $i$ in time $t+1$, assuming that the others’ ratings remain the same. We can calculate $\tilde{b}_i(\theta'_i, d(\theta'), e'_t)$ as

$$\tilde{b}_i(\theta'_i, d(\theta'), e'_t) = \sum_{kj} m_{ik} \left[ \theta_{i+1}^{(t+1)} d(\theta_{i+1}^{(t+1)}, \theta_{-i}^{(t+1)})_{kj} \right] \cdot b_i(\theta_{i+1}^{(t+1)}, \theta_{-i}^{(t+1)}).$$

(5)

where $\theta_{i+1}^{(t+1)} = \pi(\theta'_i, q_i(e'_t))$ is agent $i$’s updated rating, $d(\theta_{i+1}^{(t+1)}, \theta_{-i}^{(t+1)})$ is the new rating distribution when $i$’s rating is updated to $\theta_{i+1}^{(t+1)}$ and the others’ ratings remain the same, and $k^+_i$ is $i$’s new ranking of its new rating $\theta_{i+1}^{(t+1)}$ in the new rating distribution $d(\theta_{i+1}^{(t+1)}, \theta_{-i}^{(t+1)})$. Note that $i$ can compute $d(\theta_{i+1}^{(t+1)}, \theta_{-i}^{(t+1)})$ based on $\theta_{i+1}^{(t+1)}$ and $d(\theta')$, without knowing $\theta_{-i}^{(t+1)}$.

Each agent $i$ holds the conjecture (4) 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, and that the others’ ratings precisely reflect their review quality, namely $d(\theta')_{kj} = q_j(\theta'_j)$. Second, it conjectures that its future value is an affine function of its expected benefit. For consistency, both of the above conjectures are required to be true in the equilibrium to be defined later.

The coefficient $\alpha_i$ reflects how “optimistic” an agent is about the rating mechanism. An agent with a larger $\alpha_i$ “believes in” the rating mechanism more, because it anticipates a higher future value given the expected benefit. The coefficient $\beta'_t$ is updated in each time slot by agent $i$, such that the conjectured future value converges to the true future value in the equilibrium.

Then at each time $t$, each agent $i$ simply solves the following static problem for its optimal effort level $e'_t$:

$$e'_t = \arg \max_{e_t \in [0, \text{max}_t]} \left\{ (1 - \delta) \cdot u_i(m, \theta'_t, d(\theta'), e_t, e'_{t-1}) \right\}$$

$$+ \delta \cdot f_i(\alpha_i, \beta'_t, \theta'_t, d(\theta'), e_t).$$

(6)

Note that the others’ current effort levels $e_{-i}$ only affect the benefit term in the current payoff $u_i(m, \theta'_t, d(\theta'), e_t, e'_{t-1})$, which does not depend on $i$’s effort $e_t$ and can be considered as a constant. Hence, each agent $i$ has all the information needed to solve the above static optimization problem.

Definition 1 (Conjectural Equilibrium [17]). Given any matching rule $m$ and any rating update rule $\pi$, a conjectural equilibrium (CE) is a triple $\{\theta'_t, e'_t, \beta'_t\}_{t \in N}$ that satisfies:

- Incentive compatibility constraints: for all $i \in N'$,

$$e'_t = \arg \max_{e_t \in [0, \text{max}_t]} \left\{ (1 - \delta) \cdot u_i(m, \theta'_t, d(\theta'), e_t, e'_{t-1}) \right\}$$

$$+ \delta \cdot f_i(\alpha_i, \beta'_t, \theta'_t, d(\theta'), e_t).$$

(7)

- Stable and correct ratings: for all $i \in N'$, $\theta'_t = q_i(e'_t)$,

- Consistent conjectures: for all $i \in N'$,

$$f_i(\alpha_i, \beta'_t, \theta'_t, d(\theta'), e'_t) = u_i(m, \theta'_t, d(\theta'), e'_t).$$

(8)

In the above definition, the incentive compatibility constraints ensure that the effort level $e'_t$ is the best response of each agent $i$. In other words, it will be in agent $i$’s self-interest to choose $e'_t$. A CE also requires that each agent’s rating truly reflects its review quality at the equilibrium effort level $e'_t$, and hence each agent’s rating is stable, namely $\pi(\theta'_t, q_i(e'_t)) = \theta'_t$. Finally, a CE requires that each agent’s conjecture about its future value is correct.

There may be many 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. Aside from convergence, certain CEs are more desirable than others, as discussed in the next paragraphs.

Finally, note that there are quite a few equilibrium concepts in repeated games. One example is subgame perfect
The conjectural equilibrium appears to be the most natural and the simplest equilibrium notion, where the simplicity is in terms of the requirement on the agents rationality. For instance, SPE would require an agent to observe all the other agents actions, to have perfect recall (i.e., remember what happens in the past), to be able to predict what the other agents will do, etc. In practice, the agents have bounded rationality. Hence, the conjectural equilibrium seems to be the simplest and most natural equilibrium notion in which the agents are foresighted.

### 3.4 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 $W(q_1(e_1^i), \ldots, q_N(e_N^i))$. Then the designer problem can be defined as

$$\max_{m, \pi} W(q_1(e_1^i), \ldots, q_N(e_N^i))$$

s.t. $\{\theta_i^0, e_i^0, \beta_i^0\}_{i \in N}$ is a CE under $m, \pi$.

Note that the designer does not maximize the social welfare (i.e., the total benefit minus cost of the agents), because it is more natural from the designer’s perspective to maximize the total review quality. The designer of the peer review system may not care about the cost of reviewing; in fact, it would like to elicit more effort from the reviewers, resulting in higher-quality reviews but higher costs.

### 4 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 the best response dynamics. In this way, the designer can distinguish the true quality of the reviewers in 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:

$$e_i^t = \arg \max_{e_i \in [0, e_i^{\text{max}}]} (1 - \delta_i) \cdot u_i(m, \theta_i^t, d(\theta^t), e_i, e_i^{\text{prev}}) + \delta_i \cdot f_i(\alpha_i, \beta_i^t, \theta_i^t, d(\theta^t), e_i);$$

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

$$\beta_i^{t+1} = u_i(m, \theta_i^t, d(\theta^t), e^t) - \alpha_i \cdot \beta_i^t \cdot d(\theta^t), e_i^t).$$

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 $u_i(m, \theta_i^t, d(\theta^t), e_i, e_i^{\text{prev}})$ depends on the others’ effort levels $e_i^{\text{prev}}$, the current payoff can be separated into the benefit which depends only on the others’ effort $e_i^{\gamma}$, and the cost which depends only on agent $i$’s own effort $e_i$. Hence, when solving (10), agent $i$ can treat the benefit as a constant, and consider only the cost, which depends on its own effort level and is known to itself.

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

Next, we will provide the design guidelines on the matching rules, such that the above dynamics (10), (11), (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 agent’s expected benefit is concave and increasing in its own rating.

**Definition 2 (Desirable Matching Rules).** A matching rule $m$ is desirable, if under any rating profile $\theta$,

- each agent $i$’s (conjectured) expected benefit from the reviewing of its product, namely

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

is concave and increasing in its own rating $d(\theta)_{k_i}$;

- each agent $i$’s expected number of products to review is positive and fixed, namely

$$\sum_{k_j} m_{k_j} \cdot d(\theta)_{k_j} = M > 0.$$

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 products 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 $\theta^0$, 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), (11), (12).

**Proof.** See Appendix B, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TSNE.2018.2877578.

We illustrate the theoretical results in this paper in Fig. 1.

**Remark 1 (Uniqueness).** Theorem 1 proves that any desirable matching rule ensures the convergence to a CE. However, it is difficult to prove the convergence to a particular CE. The reasons are that there is not a single unique CE and the asymptotically reached CE (through our matching and update rules) depends on the history of the system and thus the probabilistic matching assignments and the initial ratings. In our technical analysis we prove the convergence by showing that under the updates (10), (11), (12), the
difference (in terms of $\ell_1$ norm) between two consecutive rating profiles strictly decreases over time. However, since the best responses are different under different rating profiles, the mappings from the current rating profile to the next one are different over time. Hence, the contraction mapping theorem does not apply here. In fact, in Section 5, we will show that under several desirable matching rules, there are indeed multiple CEs, and the system converges to different CEs under different initial ratings.

Remark 2 (Largest Step Size). To guarantee convergence, we require the step size $\mu$ in the rating update rule to be small enough. However, we would like the step size to be as large as possible, subject to convergence, for two reasons. First, a larger step size results in faster convergence of the ratings to the true review quality. Second, perhaps less obviously, a larger step size provides higher incentives for agents to exert high effort levels. This is because with a larger step size, the influence of the current review quality is higher on the next rating.

5 DESIGN OF MATCHING RULES

The matching rule is the critical component of our design. In this section, we will first prove that existing matching rules are inefficient. Then we propose a baseline matching rule, and analyze the properties of this baseline matching rule in detail. Finally, we propose and study two extensions.

5.1 Inefficiency of Existing Matching Rules

We show the inefficiency of existing matching rules. In existing matching rules, the matching probabilities associated with agent $i$, $m_{k_i,k_j}(d(\theta)_k,d(\theta)_j)$, do not depend on $i$’s ranking $k_i$ or its rating $d(\theta)_k$. This is true for existing matching rules in mechanism design [6], [7], [8], because there is no notion of effort and hence no rating. It is also true for existing matching rules in social norms [10], [11], [12], [13], which are uniformly random.

Proposition 1. Under any matching rule that is independent of the rating of the agent whose product will be reviewed, namely $m_{k_i,k_j}(d(\theta)_k,d(\theta)_j) = m_{k_i,k_j}(d(\theta)_{k_i},d(\theta)_{k_j})$, $\forall k_i,k_j, \theta$, there is a unique CE, in which $e^*_i = 0$ and $\theta^*_i = q_i(e^*_i) = 0$ for all $i$.

Proof. See Appendix C, available in the online supplemental material.

The above proposition shows that the matching rule that does not depend on agents’ ratings is the worst-case matching rule that results in “free-riding” by everyone. This underlines the importance of designing efficient matching rules that take into account the ratings of both the reviewers and the agents whose products are reviewed.

5.2 Design of The Baseline Matching Rule

The baseline matching rule works as follows:

1) For the agents with the same rating, match their products among themselves using any one-to-one mapping that does not match one’s product to itself.

2) For any agent $i$ with a distinct rating (i.e., no other agent has the same rating),
   a) If it has the highest rating (i.e., $k_i = 1$), match its product to a reviewer with the second highest rating with probability 1.
   b) If it has the lowest rating (i.e., $k_i = K$), match its product to a reviewer with the second lowest rating with probability $\frac{\partial(\theta)_{K-1}}{\partial(\theta)_{K-1}}$. Hence, its product gets no review with probability $1 - \frac{\partial(\theta)_{K-1}}{\partial(\theta)_{K-1}}$.
   c) If $1 < k_i < K$, match its product to its two “neighbors” with the following probabilities (which sum up to 1):

   \[
   m_{k_i,k_i-1}(d(\theta)_k,d(\theta)_{k_i-1}) = \frac{d(\theta)_{k_i} - d(\theta)_{k_i+1}}{d(\theta)_{k_i-1} - d(\theta)_{k_i+1}},
   \]
   and

   \[
   m_{k_i,k_i+1}(d(\theta)_k,d(\theta)_{k_i+1}) = \frac{d(\theta)_{k_i} - d(\theta)_{k_i-1}}{d(\theta)_{k_i-1} - d(\theta)_{k_i+1}}.
   \]

   The above matching rule is illustrated in Fig. 2. Agents with the same rating are matched to each other. For an agent with a distinct rating, it matches its product with its two nearest “neighbors” with probabilities that depend on how close its rating is to its neighbors’ ratings.

We propose this matching rule, because it has the following desirable properties:

- No agent will have to review more than 3 products. This is because any agent will at most review a product from an agent with the same rating (if there is any), and two products from its neighbors (if they have distinct ratings).

- As we will prove later, this matching rule is a desirable matching rule as defined in Definition 2.

5.2.1 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 agents at the beginning, it is reasonable to assign the same initial rating to all the agents for fairness. In this case, the following proposition tells us that we should not make the initial rating too low.

4. Although the lowest ranked agent may not be reviewed in one time slot, she will eventually be reviewed in the future. This is because the probability of getting reviewed is always positive (since an agent rating is always positive). We can consider this delay as a punishment for the lowest ranked agent.
If the agents have the same initial rating, at any point the baseline matching rule is a desirable matching rule; \( \theta \) chooses an equally matching rule for all agents is an equilibrium rating profile, and that each agent \( i \) chooses an equilibrium effort level \( e_i^* \) such that \( q_i(e_i^*) = \theta^0 \).

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

**Proof.** See Appendix D, available in the online supplemental material.

Proposition 2 implies that we should choose a high enough initial rating. In particular, when the initial rating is too low, it is optimal to choose an effort level \( e_i^* \) that satisfies \( q_i(e_i^*) = \theta^0 \). The key reason is that no agent has an incentive to reach a higher rating than the initial one, because in this case it will get a distinct rating, and get the same benefit but a higher cost 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 agent.

### 5.2.2 Convergence

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

**Definition 3 (Types).** The agents of the same type have the same normalized marginal benefit to cost ratio, defined as \( \frac{\delta_i c'_i(\cdot)}{(1-\delta_i)c_i(\cdot)} \), the same review quality function \( q_i(\cdot) \), and the same marginal benefit function \( b'_i(\cdot) \).

**Definition 4 (Ordering of Capability).** An agent \( i \) is more capable than an agent \( j \), if

\[
\frac{\delta_i c'_i(e)}{(1-\delta_i)c_i(e)} > \frac{\delta_j c'_j(e)}{(1-\delta_j)c_j(e)}, \forall e, \quad q_i(e) > q_j(e), \forall e, \quad b'_i(\theta) \geq b'_j(\theta), \forall \theta.
\]

Definition 3 defines “types” of agents, in the sense that agents of the same type will always choose the same effort level and hence get the same rating. Moreover, the type is increasing in the discount factor, the coefficient of the conjecture function, and the ratio of marginal review quality to marginal cost. Hence, the type is higher if the agent is more foresighted (i.e., weighs future payoffs more), if the agent is more “optimistic” in the sense that she believes in a higher future payoff given the same benefit, and if the agent’s effort results in higher review quality and lower cost. Hence, this definition of the type is a reasonable measure of an agent’s capability. Definition 4 gives an ordering of agents in terms of their “capability”. We will prove that a more capable agent indeed gets a higher rating.

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

**Assumption 2 (Large Population).** There is more than one agent of each type.

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

Note that although there is no agent with distinct ratings under Assumption 2, the matching rule still needs to specify the matching outcomes in all the scenarios, including when an agent has a distinct rating. This specification is necessary for the agents to compute their best-response effort levels, and will affect the equilibrium rating profile.

**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 \( \mu > 0 \) such that under any small step size \( \mu \in (0, \mu] \) in the rating update rule, the system will converge to a CE through the best response dynamics (10), (11), (12);
- if the agents have the same initial rating, at any point in the best response dynamics (10)–(12), more capable agents will always have no lower ratings than less capable agents.

**Proof.** See Appendix E, available in the online supplemental material.

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. That is, a more capable agent never has lower ratings than a less capable agent at any point in the best response dynamics. This means that the rating mechanism can distinguish the agents of different types, and rank them in the correct order. Note that more capable agents produce reviews of high enough quality (that result in higher ratings) in their self-interest, as a result of maximizing their own payoffs; they are not obliged to do so by the designer.

| Rating | Author | Matching | Reviewer | Rating |
|--------|--------|----------|----------|--------|
| 1      | 0.5    |          | 0.1      | 1      |
| 0.9    | 0.5    |          | 0.2      | 0.9    |
| 0.8    | 0.5    | 0.2      | 0.33     | 0.8    |
| 0.4    | 0.5    | 0.33     | 0.1      | 0.4    |
| 0.2    | 0.67   | 0.1      | 0.1      | 0.2    |
| 0.1    | 0.33   | 0.1      | 0.1      | 0.1    |

Fig. 2. Illustration of the baseline matching rule with seven agents. The highest-ranking agent’s product is matched to the reviewer with the second highest rating. The lowest-ranking agent’s product is matched to the reviewer with the second lowest rating with probability 0.5. The two agents 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.
5.3 Two Classes of Extended Matching Rules

Previously, we have focused on the baseline matching rule. The baseline matching rule is able to incentivize the agents to exert high effort levels by increasing the benefit obtained by an agent when its rating increases. Now we extend the baseline rule in two different ways, both of which result 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 asymmetric probabilities for matching an agent with a distinct rating to its higher and lower neighbors. In particular, the asymmetric matching rule is parametrized by $\gamma$ such that any agent $i$ with a distinct rating and with $k_i \in [2, K-1]$ is matched to its neighbors with the following probabilities:

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

and

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

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

We illustrate this asymmetric extension in Fig. 3. We can see that when $\gamma > 0$ ($\gamma < 0$), the resulting matching rule rewards (punishes) the agent by increasing its probability of being matched to the higher-rating (lower-rating) neighbor. When $\gamma = 0$, the asymmetric matching rule reduces to the baseline matching rule.

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

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

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

and

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

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

We refer to this extension as long-range matching rule; see Fig. 4 for an illustration.

We can see that the parameters $\gamma_r$ and $\gamma_p$ reflect to what extent the agents 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 agent is rewarded (punished) by being matched to a reviewer with the next higher (lower) rating.

We summarize the key differences among the baseline matching rule and its extensions in Fig. 5.

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.

![Fig. 3. An illustration of the first asymmetric extension of the baseline matching rule. We only show the matching of the agent with rating 0.8, who is matched to its two nearest neighbors with different probabilities than in the baseline rule.](image)

![Fig. 4. An illustration of the second long-range extension of the baseline matching rule. We show the matching of only one agent with rating 0.8, who is matched to its four neighbors, instead of two nearest neighbors as in the baseline rule.](image)

![Fig. 5. Comparison of the baseline matching rule and its two extensions. We illustrate how an agent with a certain rating (the red dot) is matched. In the baseline rule, it is matched to its nearest neighbors with equal probability. In the first asymmetric extension, although the distances between the ratings remain the same, the matching probabilities are slightly changed. In the second long-range extension, it can be matched to neighbors with even higher or lower ratings.](image)
We say that a matching rule \( m' \) is “better” than another matching rule \( m \), if for any equilibrium rating profile \( \theta^* \) under \( m \), we can find an equilibrium rating profile \( \theta'^* \) under \( m' \) that satisfies \( \theta'^* > \theta^* \).

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 asymmetric extension, there exists a \( \gamma > 0 \) (i.e., extra reward) under which the asymmetric matching rule is strictly better than the baseline matching rule.
- In the second long-range extension, there exists \( \gamma_r = 0 \) and \( \gamma_p > 0 \) (i.e., extra punishment) under which the long-range matching rule is strictly better than the baseline matching rule.

**Proof.** See Appendix F, available in the online supplemental material.

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 two extended matching rules. The technical reason is that we want to increase the marginal expected benefit when an agent’s rating is changed, in order to give more incentive for them to exert high effort levels. The main message delivered by our result is that we should carefully design the matching rule based on the way we reward and punish.

## Illustrative Results

We consider a course in a MOOC platform. We choose the number of agents to be 1000, a typical number in MOOCs [2], [3], [4], [5], where there are 10 types of agents with 100 agents of each type. All the agents have the same patience \( \delta_i = 0.8 \). We interpret the review quality as the accuracy of the grading and the effort as the time spent in grading. Then linear quality functions are observed in practice (see, e.g., Fig. 5a and 5b in [18]). Therefore, we set different quality functions \( q_i(e_i) = p_i \cdot e_i \) for different types of agents with \( p_i = 0.2, 0.4, 0.6, \ldots, 2.0 \). They also have different \( \alpha_i = 0.2, 0.4, 0.6, \ldots, 2.0 \) in the conjecture functions. Finally, quadratic benefit and cost functions are commonly used in game theory literature. Therefore, we set the same cost function \( c_i(e_i) = e_i^2 \), and the same benefit function \( b_i(\theta) = -\theta^2 + 2\theta \) for all the agents.

### 6.1 Impact of Step Sizes in Rating Update

We show the best response dynamics under step sizes \( \mu = 0.3 \) and \( \mu = 0.9 \) in Figs. 6 and 7, respectively. Note that we only show the ratings of agents of types 1, 3, 5, 7, 9. We first observe that under one-shot or exogenous matching rules, the reviewers exert 0 efforts and get 0 ratings all the time. Hence, the proposed endogenous matching greatly improves the performance of the system.

Second, we can see that under a larger step size, the agents’ equilibrium ratings are higher, indicating higher equilibrium effort levels and higher equilibrium review quality. This is consistent with our intuition in Remark 2. Recall from Remark 2 that theoretically the convergence is guaranteed by a small enough step size. However, as we have seen in the simulation, the ratings still converge under a large step size of \( \mu = 0.9 \). In this case, all the agents’ ratings converge in three iterations. This suggests that the proposed scheme is applicable in practical scenarios that require fast convergence.

### 6.2 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 1, we evaluate the first asymmetric extension of matching rules under different parameters \( \gamma \). We can see that in our setting, the optimal \( \gamma \) should be 0.1, which results

| \( \gamma \) | -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 2
Equilibrium Review Quality and Social Welfare under the Second Long-Range Extension of Matching Rules

| (y_r, y_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 |

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 fact, in terms of social welfare, the optimal $y$ is $-0.05$, which results in lower review quality and thus lower cost. How the review quality and the social welfare are aligned depends on the benefit and cost functions.

In Table 2, we evaluate the second long-range extension of matching rules under different parameters $y_r$ and $y_p$. We can see that the optimal sum review quality is achieved when $y_r = 0$ and $y_p = 1$, 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 agents to exert high effort. Again, such a matching rule does not result in the optimal social welfare. The optimal social welfare is achieved when $y_r = 0.5$ and $y_p = 0.5$, where the agents are also rewarded.

7 Conclusion

In this work, we proposed the first rating and repeated endogenous matching mechanisms to address the adverse selection and moral hazard problems simultaneously in peer review. Our proposed rating and matching mechanisms are easy to implement, require no knowledge of agents’ private information, and ensure the convergence to an equilibrium in which the agents get their review quality revealed by their ratings and are incentivized to produce high-quality reviews. We thoroughly studied the design of matching rules, in terms of the initial ratings, the requirements for convergence, and the equilibrium ratings and review quality. We also studied extensions to different classes of matching rules, and proved the optimality of different rewarding/punishing mechanisms under different matching rules.

In future work, we will investigate the effect of inaccurate and possibly biased reports about the review quality, as well as more detailed modeling frameworks including multiple agents and reviewers associated with each product. An intriguing problem is the quest for the best matching rule maximizing the review quality.

Acknowledgments

Y. Xiao and M. van der Schaar were supported by NSF CCF Grant No. 1218136. The work was done when they were with Department of Electrical Engineering, UCLA, CA.

References

[1] Y. Xiao, F. Dörfler, and M. van der Schaar, “Rating and matching in peer review systems,” in Proc. 52nd Annu. Allerton Conf. Commun. Control. Comput. (Allerton), 2014, pp. 54–61.
[2] J. Daniel, “Making sense of MOOCs: Musings in a maze of myth, paradox and possibility,” J. Interactive Media Edu., vol. 3, 2012, Art. no. 18.
[3] C. Piech, J. Huang, Z. Chen, C. Do, A. Ng, and D. Koller, “Tuned models of peer assessment in MOOCs,” in Proc. Int. Conf. Educational Data Mining, 2013, pp. 1–8.
[4] C. G. Brinton, M. Chiang, S. Jain, H. Lam, Z. Liu, and F. M. F. Wong, “Learning about social learning in MOOCs: From statistical analysis to generative model,” IEEE Trans. Learn. Technol., vol. 7, no. 4, pp. 346–359, Oct.-Dec. 2014.
[5] N. B. Shah, J. K. Bradley, A. Parekh, M. Wainwright, and K. Ramchandran, “A case for ordinal peer evaluation in MOOCs,” NIPS Workshop Data Driven Educ., 2013, pp. 1–8.
[6] D. Gale and L. S. Shapley, “College admissions and the stability of marriage,” Amer. Math. Monthly, vol. 69, pp. 9–14, 1962.
[7] A. Abdulkadiroğlu and T. Sonmez, “School choice: A mechanism design approach,” Amer. Econ. Rev., vol. 93, no. 3, pp. 729–747, 2003.
[8] A. E. Roth, T. Sonmez, and M. UtkuÜver, “Pairwise kidney exchange,” J. Econ. Theory, vol. 125, no. 2, pp. 151–188, 2005.
[9] N. Arnosti, R. Johari, and Y. Kanoria, “Managing congestion in decentralized matching markets,” 2014. [Online]. Available at SSRN: http://dx.doi.org/10.2139/ssrn.2427960
[10] M. Kandori, “Social norms and community enforcement,” Rev. Econ. Studies, vol. 59, no. 1, pp. 63–80, 1992.
[11] Y. Zhang and M. van der Schaar, “Peer-to-peer multimedia sharing based on social norms,” Elsevier J. Signal Process.: Image Commun. Special Issue on Adv. Video Streaming P2P Netw., vol. 27, no. 5, pp. 383–400, May 2012.
[12] Y. Zhang and M. van der Schaar, “Rating protocols for online communities,” ACM Trans. Econ. Comput., vol. 2, 2013, Art. no. 4.
[13] Y. Xiao and M. van der Schaar, “Socially-optimal design of service exchange platforms with imperfect monitoring,” ACM Trans. Economics Comput., vol. 3, no. 4, Jul. 2015.
[14] D. Kurokawa, O. Lev, J. Morgenstern, and A. D. Procaccia, “Impartial Peer Review,” Working Paper, 2014. [Online]. Available: http://www.cs.cmu.edu/arielpro/papers/impartial.pdf
[15] A. Spalvieri, S. Mandelli, M. Magarini, and G. Bianchi, “Weighting peer reviewers,” in Proc. Int. Conf. Privacy Secur. Trust, 2014, pp. 414–419.
[16] J. Prüfer and D. Zetland, David, “An auction market for journal articles”, Public Choice, vol. 145, no. 3/4, pp. 379–403, 2010.
[17] F. H. Hahn, “Exercises in conjectural equilibria,” Scandinavian J. Econ., vol. 79, no. 2, pp. 210–226, 1977.
[18] C.-J. Ho, A. Slinkins, S. Suri, and J. W. Vaughan, “Incentivizing high quality crowdwork,” in Proc. Int. World Wide Web Conf., 2015, pp. 417–429.

Yuanzhang Xiao received the BE and ME degrees in electronic engineering from Tsinghua University, in 2006 and 2009, respectively and the PhD degree in electrical engineering from UCLA, in 2014. He is an assistant professor with the University of Hawaii at Manoa. He was a postdoctoral fellow with Northwestern University from 2015 to 2017. His research interests include game theory, mechanism design, and optimization, with applications in socio-technological networks, smart grids, and wireless communication. He is a member of the IEEE.
Florian Dorfler is an Assistant Professor at the Automatic Control Laboratory at ETH Zürich. He received his Ph.D. degree in Mechanical Engineering from the University of California at Santa Barbara in 2013, and a Diplom degree in Engineering Cybernetics from the University of Stuttgart in 2008. From 2013 to 2014 he was an Assistant Professor at the University of California Los Angeles. His primary research interests are centered around distributed control, complex networks, and cyber–physical systems currently with applications in energy systems and smart grids. His students were winners or finalists for Best Student Paper awards at the 2013 European Control Conference, the 2016 American Control Conference, and the 2017 PES PowerTech Conference. His articles received the 2010 ACC Student Best Paper Award, the 2011 O. Hugo Schuck Best Paper Award, the 2012-2014 Automatica Best Paper Award, and the 2016 IEEE Circuits and Systems Guillemin-Cauer Best Paper Award. He is a recipient of the 2009 Regents Special International Fellowship, the 2011 Peter J. Frenkel Foundation Fellowship, and the 2015 UCSB ME Best PhD award.

Mihaela van der Schaar (F’09) is a chancellor's professor of electrical engineering with the University of California, Los Angeles. She was a distinguished lecturer of the IEEE Communications Society for 2011-2012 and an editor in chief of the IEEE Transactions on Multimedia (2011-2013). Her research interests include networks, game theory, network science, social networks, complex systems, machine learning, online learning, stream mining and Big data, and medical informatics. She received an NSF CAREER Award (2004), the Best Paper Award from the IEEE Transactions on Circuits and Systems for Video Technology (2005), the Okawa Foundation Award (2006), the IBM Faculty Award (2005, 2007, and 2008), the Most Cited Paper Award from the EURASIP: Image Communications Journal (2006), the GameNets Conference Best Paper Award (2011), and the 2011 IEEE Circuits and Systems Society Darlington Award Best Paper Award. She received 33 granted US patents. She is also the founding and managing director of the UCLA Center for Engineering Economics, Learning, and Networks (see http://netecon.ee.ucla.edu). For more information about her research, visit: http://medianetlab.ee.ucla.edu/

For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.