

# Socially-Optimal Design of Service Exchange Platforms with Imperfect Monitoring

Yuanzhang Xiao and Mihaela van der Schaar

## Abstract

We study the design of service exchange platforms in which long-lived anonymous users exchange services with each other. The users are randomly matched into pairs of clients and servers repeatedly, and each server can choose whether to provide high-quality or low-quality services to the client with whom it is matched. Since the users are anonymous and incur high costs (e.g. exert high effort) in providing high-quality services, it is crucial that the platform incentivizes users to provide high-quality services. Rating mechanisms have been shown to work effectively as incentive schemes in such platforms. A rating mechanism labels each user by a rating, which summarizes the user’s past behaviors, recommends a desirable behavior to each server (e.g., provide higher-quality services for clients with higher ratings), and updates each server’s rating based on the recommendation and its client’s report on the service quality. Based on this recommendation, a low-rating user is less likely to obtain high-quality services, thereby providing users with incentives to obtain high ratings by providing high-quality services.

However, if monitoring or reporting is imperfect – clients do not perfectly assess the quality or the reports are lost – a user’s rating may not be updated correctly. In the presence of such errors, existing rating mechanisms cannot achieve the social optimum. In this paper, we propose the first rating mechanism that does achieve the social optimum, even in the presence of monitoring or reporting errors. On one hand, the socially-optimal rating mechanism needs to be complicated enough, because the optimal recommended behavior depends not only on the current rating distribution, but also (necessarily) on the history of past rating distributions in the platform. On the other hand, we prove that the social optimum can be achieved by “simple” rating mechanisms that use *binary* rating labels and a *small* set of (three) recommended behaviors. We provide design guidelines of socially-optimal rating mechanisms, and a low-complexity online algorithm for the rating mechanism to determine the optimal recommended behavior.

---

We would like to thank Yu Zhang for very insightful initial discussions on rating mechanisms, and Prof. William Zame and Simpson Zhang (Department of Economics, UCLA) for their helpful comments that lead to the improvement of this paper.

## I. INTRODUCTION

Service exchange platforms have proliferated as the medium that allows the users to exchange services valuable to each other. For instance, emerging new service exchange platforms include crowdsourcing systems (e.g. in Amazon Mechanical Turk and CrowdSource) in which the users exchange labor [1][2], online question-and-answer websites (e.g. in Yahoo! Answers and Quora) in which the users exchange knowledge [2], peer-to-peer (P2P) networks in which the users exchange files/packets [3][4][5], and online trading platforms (e.g. eBay) where the users exchange goods [6]. In a typical service exchange platform, a user plays a dual role: as a *client*, who requests services, and as a *server*, who chooses to provide high-quality or low-quality services. Common features of many service exchange platforms are: the user population is large and users are anonymous. In other words, each user interacts with a randomly-matched partner without knowing its partner’s identity (However, the platform does know the identify of the interacting users.). The absence of a fixed partner and the anonymity of the users create incentive problems – namely the users tend to “free-ride” (i.e., receive high-quality services from others as a client, while providing low-quality services as a server). In addition, a user generally may not be able to perfectly monitor<sup>1</sup> its partner’s action, which makes it even harder to incentivize the users to provide high-quality services.

An important class of incentive mechanisms for service exchange platforms are the rating mechanisms<sup>2</sup> [2]–[13], in which each user is labeled with a rating based on its past behaviors in the system. A rating mechanism consists of a rating update rule and a recommended strategy<sup>3</sup>. The recommended strategy specifies what is the desirable behavior under the current system state (e.g. the current rating profile of the users or the current rating distribution). For example, the rating mechanism may recommend providing high-quality services for all the users when the

---

<sup>1</sup>The monitoring discussed throughout this paper is a user’s observation on its current partner’s actions. Each user knows nothing about the ongoing interactions among the other pairs of users.

<sup>2</sup>Note that the rating mechanisms studied in this paper focus on dealing with moral hazard problems, namely the server’s quality of service is not perfectly observable. They are different from the rating mechanisms dealing with adverse selection problems, namely the problems of identifying the users’ types. See [6, Sec. I] for detailed discussions on the above two classes of rating mechanisms.

<sup>3</sup>Different terminologies have been used in the existing literature. For example, [6][7] used “reputation” for “rating”, and [7] used “social norm” for “recommended strategy”.

majority of users have high ratings, while recommending to provide high-quality services only to high-rating users when the majority have low ratings. Then, based on each client’s report on the quality of service, the rating mechanism revises each server’s rating according to the rating update rule. Generally speaking, the ratings of the users who comply with (resp. deviate from) the recommended behaviors go up (resp. down). Hence, each user’s rating summarizes its past behavior in the system. By keeping track of all the users’ ratings and recommending them to reward (resp. punish) the users with high (resp. low) ratings, the rating mechanism gives incentives to the users to obtain high ratings by rewarding them indirectly, through recommending other users to provide them with high-quality services.

Existing rating mechanisms have been shown to work well when monitoring and reporting are perfect. However, when monitoring and reporting are subject to errors, existing rating mechanisms cannot achieve the social optimum [2]–[13]. The errors, which are often encountered in practice, may arise either from the client’s own incapability of accurate assessment (for instance, the client, who wants to translate some sentences into a foreign language, cannot accurately evaluate the server’s translation), or from some system errors (for example, the client’s report on the server’s service quality is missing due to network errors)<sup>4</sup>. In the presence of errors, the server’s rating may be wrongly updated. Hence, even if the users follow the recommended desirable behavior, the platform may still fall into some “bad” states in which many users have low ratings due to erroneous rating updates. In these bad states, the users with low ratings receive low-quality services, resulting in large performance loss compared to the social optimum. This performance loss in the bad states is the major reason for the inefficiency of the existing rating mechanisms.

In this paper, we propose the first rating mechanisms that can achieve the social optimum even under imperfect monitoring. A key feature of the proposed rating mechanism is the *nonstationary* recommended strategy, which recommends different behaviors under the same system state, depending on when this state occurs (for example, the rating mechanism may not always recommend punishing users with low ratings in the bad states). Note, importantly, that the

---

<sup>4</sup>Note that the errors in this paper are not caused by the strategic behaviors of the users. In other words, the clients report the service quality truthfully, and do not misreport intentionally to manipulate the rating mechanism for their own interests. If the clients may report strategically, the mechanism can let the platform to assess the service quality (still, with errors) to avoid strategic reporting.

rating mechanism does not just randomize over different behaviors with a fixed probability in a state. Instead, it recommends different behaviors in the current state based on the history of past states. We design the recommended strategy carefully, such that the punishments happen frequently enough to provide sufficient incentives for the users, but not too frequently to reduce the performance loss incurred in the bad states. The more patient the users are (i.e. the larger discount factor they have), the less frequent are the punishments. As a result, the designed rating mechanism can asymptotically achieve the social optimum as the users become increasingly patient (i.e. as the discount factor approaches 1). This is in contrast with the existing rating mechanisms with *stationary* recommended strategies, whose performance loss does not vanish even as the users' patience increases. Another key feature of the proposed rating mechanism is the use of differential punishments that punish users with different ratings differently. In Section IV, we show that the absence of any one of these two features in our mechanism will result in performance loss that does not vanish even when the users are arbitrarily patient.

We prove that the social optimum can be achieved by simple rating mechanisms, which assign *binary* ratings to the users and recommend a small set of *three* recommended behaviors. We provide design guidelines of the rating update rules in socially-optimal rating mechanisms, and a low-complexity online algorithm to construct the nonstationary recommended strategies. The algorithm essentially solves a system of two linear equations with two variables in each period, and can be implemented with a memory of a fixed size (although by the definition of nonstationary strategies, it appears that we may need a memory growing with time to store the history of past states), because we can appropriately summarize the history of past states (by the solution to the above linear equations).

The rest of the paper is organized as follows. In Section II, we discuss the differences between our work and related works. In Section III, we describe the model of service exchange systems with rating mechanisms. Then we design the optimal rating mechanisms in Section V. Simulation results in Section VI demonstrate the performance improvement of the proposed rating mechanism. Finally, Section VII concludes the paper.

TABLE I. RELATED WORKS ON RATING PROTOCOLS.

	Rating update error	Recommended strategy	Discount factor	Performance loss
[2][3]	$\rightarrow 0$	Stationary	$< 1$	Yes
[4][5]	$> 0$	Stationary	$< 1$	Yes
[6]	$> 0$	Stationary/Nonstationary	$< 1$	Yes
[7]–[12]	$= 0$	Stationary	$\rightarrow 1$	Yes
[13]	$\rightarrow 0$	Stationary	$\rightarrow 1$	Yes
This work	$> 0$	Nonstationary	$< 1$	No

## II. RELATED WORKS

### A. Related Works on Rating Protocols

Rating mechanisms were originally proposed by [7] for a large anonymous society, in which users are repeatedly randomly matched to play the Prisoners’ dilemma game. Assuming perfect monitoring, [7] proposed a simple rating mechanism that can achieve the social optimum: any user who has defected will be assigned with the lowest rating forever and will be punished by its future partners. Subsequent research has been focusing on extending the results to more general games (see [8][9][10][12]), or on discovering alternative mechanisms (for example, [11] showed that cooperation can be sustained if each user can observe its partner’s past actions). However, all these works assumed perfect monitoring and were aimed at dealing with the incentive problems caused by the anonymity of users and the lack of fixed partnership; they did not study the impact of imperfect monitoring. Under imperfect observation/reporting, the system will collapse under their rating mechanisms because all the users will eventually end up with having low ratings forever due to errors.

Some works [2][3][13] assumed imperfect monitoring, but focused on the limit case when the monitoring tends to be perfect. The conclusion of these works is that the social optimum can be achieved in the limit case when the monitoring becomes “almost perfect” (i.e., when the rating update error goes to zero).

Only a few works [4]–[6] analyzed rating mechanisms under imperfect monitoring with *fixed nonzero* monitoring errors. For a variety of rating mechanisms studied in [4]–[6], the performance loss with respect to the social optimum is quantified in terms of the rating update error. These results confirm that existing rating mechanisms suffer from (severe) performance loss under rating update errors. Note that the model in [6] is fundamentally different than ours. In [6], there

is only a single long-lived seller (server), while all the buyers (clients) are short-lived. Under this model, it is shown in [6] that the rating mechanism is bounded away from social optimum even when nonstationary strategies are used. In contrast, we show that under our model with long-lived servers and clients, we can achieve the social optimum by nonstationary strategies with differential punishments. In the following, we discuss the intuitions of how to achieve the social optimum under our model.

There are two sources of inefficiency. One source of inefficiency comes from the stationary recommended strategies, which recommends the same behavior under the same state [2]–[5][7]–[13]. As we have discussed earlier, the inefficiency of the existing rating mechanisms comes from the punishments triggered in the “bad” states. Specifically, to give incentives for the users to provide high-quality services, the rating mechanism *must* punish the low-rating users under certain rating distributions (i.e. under certain “bad” states). When the users are punished (i.e. they are provided with low-quality services), the average payoffs in these states are far below the social optimum. In the presence of rating update errors, the bad states happen with a probability bounded above zero (the lower bound depends only on the rating update error). As a result, the low payoffs occur with a frequency bounded above zero, which incurs an efficiency loss that cannot vanish unless the rating update error goes to zero.

Another source of inefficiency is the lack of differential punishments. As will be proved in Section IV, the rating mechanisms with no differential punishment have performance loss even when nonstationary recommended strategies are used.

This paper is the first to propose a class of rating mechanisms that achieve the social optimum even when update errors do not tend to zero. Our mechanisms rely on (explicitly-constructed) *nonstationary* strategies with differential punishments. The key intuitions of why the proposed mechanism achieves social optimum are as follows. First, nonstationary strategies punish the users in the bad states only when necessary, depending on the history of past states. In this way, nonstationary strategies can lower the frequency of punishment in the bad states to a level just enough to provide sufficient incentives for the users to provide high-quality services. In addition, differential punishment further reduces the loss in social welfare by transferring payoffs from low-rating users to high-rating users, instead of lowering everyone’s payoff with non-differential punishment.

In Table I, we compare the proposed work with existing rating mechanisms.

## B. Related Works in Game Theory Literature

Our results are related to folk theorem results for repeated games [17] and stochastic games [18]. However, these existing folk theorem results [17][18] cannot be directly applied to our model. First, the results in [17] are derived for repeated games, in which every stage game is the same. Our system is modeled as a stochastic game, in which the stage games may be different because of the rating distributions.

Second, there do exist folk theorems for stochastic games [18], but they also do not apply to our model. The folk theorems [18] apply to standard stochastic games, in which the state must satisfy the following properties: 1) the state, together with the plan profile, uniquely determines the stage-game payoff, and 2) the state is known to all the users. In our model, since each user's stage game payoff depends on its own rating, each user's rating must be included in the state and be known to all the users. In other words, if we model the system as a standard stochastic game in order to apply the folk theorems, we need to define the state as the rating profile of all the users (not just the rating distribution). Then, the folk theorem states that the social optimum can be asymptotically achieved by strategies that depend on the history of rating profiles. However, in our model, the players do not know the full rating profile, but only know the rating distribution. Hence, the strategy can use only the information of rating distributions.<sup>5</sup> Whether such strategies can achieve the social optimum is not known according to the folk theorems; we need to prove the existence of socially optimal strategies that use only the information of rating distributions.

In addition, our results are fundamentally different from the folk theorem results [17][18] in nature. First, [17][18] focus on the limit case when the discount factor goes to one, which is not realistic because the users are not sufficiently patient. More importantly, the results in [17][18] are not constructive. They focus on *what* payoff profiles are achievable, but cannot show *how* to achieve those payoff profiles. They do not determine a lower bound on discount factors that admit equilibrium strategy profiles yielding the target payoff profile, and hence cannot construct equilibrium strategy profiles. By contrast, we do determine a lower bound on discount factors that admit equilibrium strategy profiles yielding the target payoff profile, and do construct equilibrium

---

<sup>5</sup>We insist on restricting to strategies that depend only on the history of rating distributions because in practice, 1) the platform may not publish the full rating profile due to informational and privacy constraints, and 2) even if the platform does publish such information, it is impractical to assume that the users can keep track of it.

TABLE II. RELATED MATHEMATICAL FRAMEWORKS.

	Standard MDP	Extended MDP [2][3][14][15]	Self-generating sets [16]–[18]	This work
# of users	Single	Multiple	Multiple	Multiple
Value function	Single-valued	Single-valued	Set-valued	Set-valued
Incentive constraints	No	Yes	Yes	Yes
Strategies	Stationary	Stationary	Nonstationary	Nonstationary
Discount factor	$< 1$	$< 1$	$\rightarrow 1$	$< 1$
Constructive	Yes	Yes	No	Yes

strategy profiles.

### C. Related Mathematical Frameworks

Rating mechanisms with stationary recommended strategies can be designed by extending Markov decision processes (MDPs) in two important and non-trivial ways [2][3][14][15]: 1) since there are multiple users, the value of each state is a vector of all the users’ values, instead of a scalar in standard MDPs, and 2) the incentive compatibility constraints of self-interested users need to be fulfilled (e.g., the values of “good” states, in which most users have high ratings, should be sufficiently larger than those of “bad” states, such that users are incentivized to obtain high ratings), while standard MDPs do not impose such constraints.

In this paper, we make a significant step forward with respect to the state-of-the-art rating mechanisms with stationary strategies: we design rating mechanisms where the recommended strategies can be nonstationary. The proposed design leads to significant performance improvements, but is also significantly more challenging from a theoretical perspective. The key challenge is that nonstationary strategies may choose different actions under the same state, resulting in possibly different current payoffs in the same state. Hence, the value function under nonstationary strategies are *set-valued*, which significantly complicates the analysis, compared to *single-valued* value functions under stationary strategies<sup>6</sup>.

The mathematical framework of analyzing nonstationary strategies with set-valued value functions was proposed as a theory of self-generating sets in [16]. It was widely used in game theory

---

<sup>6</sup>In randomized stationary strategies, although different actions may be taken in the same state *after randomization*, the probability of actions chosen is fixed. In the Bellman equation, we need to use the *expected payoff before randomization* which is fixed in the same state, instead of the realized payoffs after randomization. Hence, the value function is still single-valued.



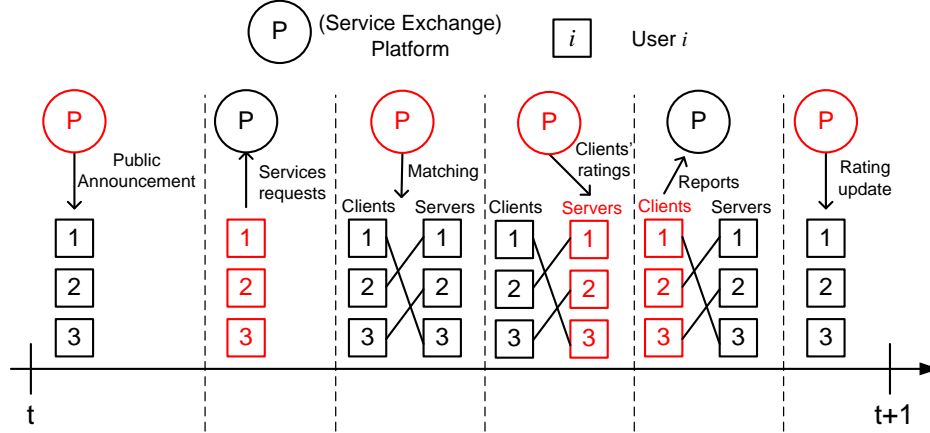


Fig. 1. Illustration of the rating mechanism in one period.

to prove folk theorems in repeated games [17] and stochastic games [18]. We have discussed our differences from the folk theorem results [17][18] in the previous subsection.

In Table II, we compare our work with existing mathematical frameworks.

### III. SYSTEM MODEL AND PROBLEM FORMULATION

#### A. System Model

1) *The Rating Mechanism:* We consider a service exchange platform with a set of  $N$  users, denoted by  $\mathcal{N} = \{1, \dots, N\}$ . Each user can provide some services (e.g. data in P2P networks, labor in Amazon Mechanic Turk) valuable to the other users. The rating mechanism assigns each user  $i$  a binary label  $\theta_i \in \Theta \triangleq \{0, 1\}$ , and keep record of the *rating profile*  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$ . Since the users usually stay in the platform for a long period of time, we divide time into periods indexed by  $t = 0, 1, 2, \dots$ . In each period, the rating mechanism operates as illustrated in Fig. 1, which can be roughly described as follows:

- Each user requests services as a *client*.
- Each user, as a *server*, is matched to another user (its client) based on a matching rule.
- Each server chooses to provide high-quality or low-quality services.
- Each client reports its assessment of the service quality to the rating mechanism, who will update the server's rating based on the report.

Next, we describe the key components in the rating mechanism in details.

*Public announcement:* At the beginning of each period, the platform makes public announcement to the users. The public announcement includes the rating distribution and the recommended plan in this period. The *rating distribution* indicates how many users have rating 1 and rating 0, respectively. Denote the rating distribution by  $\mathbf{s}(\boldsymbol{\theta}) = (s_0(\boldsymbol{\theta}), s_1(\boldsymbol{\theta}))$ , where  $s_1(\boldsymbol{\theta}) = \sum_{i \in \mathcal{N}} \theta_i$  is the number of users with rating 1, and  $s_0(\boldsymbol{\theta}) = N - s_1(\boldsymbol{\theta})$  is the number of users with rating 0. Denote the set of all possible rating distributions by  $S$ . Note that the platform does not disclose the rating profile  $\boldsymbol{\theta}$  for privacy concerns. The platform also recommends a desired behavior in this period, called *recommended plan*. The recommended plan is a contingent plan of which service quality the server should choose based on its own rating and its client's rating. Formally, the recommended plan, denoted by  $\alpha_0$ , is a mapping  $\alpha_0 : \Theta \times \Theta \rightarrow \{0, 1\}$ , where 0 and 1 represent “low-quality service” and “high-quality service”, respectively. Then  $\alpha_0(\theta_c, \theta_s)$  denotes the recommended service quality for a server with rating  $\theta_s$  when it is matched to a client with rating  $\theta_c$ . We write the set of recommended plans as  $\mathcal{A} = \{\alpha | \alpha : \Theta \times \Theta \rightarrow \{0, 1\}\}$ . We are particularly interested in the following three plans. The plan  $\alpha^a$  is the *altruistic* plan:

$$\alpha^a(\theta_c, \theta_s) = 1, \forall \theta_c, \theta_s \in \{0, 1\}, \quad (1)$$

where the server provides high-quality service regardless of its own and its client's ratings. The plan  $\alpha^f$  is the *fair* plan:

$$\alpha^f(\theta_c, \theta_s) = \begin{cases} 0 & \theta_s > \theta_c \\ 1 & \theta_s \leq \theta_c \end{cases}, \quad (2)$$

where the server provides high-quality service when its client has higher or equal ratings. The plan  $\alpha^s$  is the *selfish* plan:

$$\alpha^s(\theta_c, \theta_s) = 0, \forall \theta_c, \theta_s \in \{0, 1\}, \quad (3)$$

where the server provides low-quality service regardless of its own and its client's ratings. Note that we can consider the selfish plan as a non-differential punishment in which everyone receives low-quality services, and consider the fair plan as a differential punishment in which users with different ratings receive different services.

*Service requests:* The platform receives service requests from the users. We assume that there is no cost in requesting services, and that each user always have demands for services. Hence, all the users will request services.

*Matching:* The platform matches each user  $i$ , as a client, to another user  $m(i)$  who will serve  $i$ , where  $m$  is a matching  $m : \mathcal{N} \rightarrow \mathcal{N}$ . Since the platform cannot match a user to itself, we write the set of all possible matchings as  $M = \{m : m \text{ bijective, } m(i) \neq i, \forall i \in \mathcal{N}\}$ . The mechanism defines a random matching rule, which is a probability distribution  $\mu$  on the set of all possible matchings  $M$ . In this paper, we focus on the uniformly random matching rule, which chooses every possible matching  $m \in M$  with the same probability. The analysis can be easily generalized to the cases with non-uniformly random matching rules, as long as the matching rules do not distinguish users with the same rating.

*Clients' ratings:* The platform will inform each server of its client's rating, such that each server can choose its service quality based on its own and its client's ratings.

*Reports:* After the servers serve their clients, the platform elicits reports from the clients about their service quality. However, the report is *inaccurate*, either by the client's incapability of accurate assessment (for instance, the client, who wants to translate some sentences into a foreign language, cannot accurately evaluate the server's translation) or by some system error (for example, the data/file sent by the server is missing due to network errors). We characterize the erroneous report by a mapping  $R : \{0, 1\} \rightarrow \Delta(\{0, 1\})$ , where  $\Delta(\{0, 1\})$  is the probability distribution over  $\{0, 1\}$ . For example,  $R(1|q)$  is the probability that the client reports "high quality" given the server's actual service quality  $q$ . In this paper, we focus on reports of the following form

$$R(r|q) = \begin{cases} 1 - \varepsilon, & r = q \\ \varepsilon, & r \neq q \end{cases}, \quad \forall r, q \in \{0, 1\}, \quad (4)$$

where  $\varepsilon \in [0, 0.5)$  is the report error probability.<sup>7</sup> Note, however, that reporting is not strategic: the client reports truthfully, but with errors. If the clients report strategically, the mechanism can let the platform to assess the service quality (still, with errors) to avoid strategic reporting. For simplicity, we assume that the report error is symmetric, in the sense that reporting high and low qualities have the same error probability. Extension to asymmetric report errors is straightforward.

---

<sup>7</sup>We confine the report error probability  $\varepsilon$  to be smaller than 0.5. If the error probability  $\varepsilon$  is 0.5, the report contains no useful information about the service quality. If the error probability is larger than 0.5, the rating mechanism can use the opposite of the report as an indication of the service quality, which is equivalent to the case with the error probability smaller than 0.5.

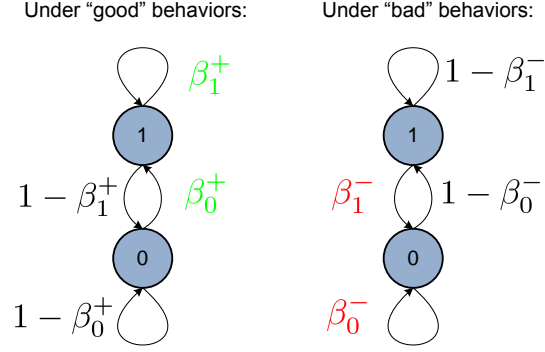


Fig. 2. Illustration of the rating update rule. The circle denotes the rating, and the arrow denotes the rating update with corresponding probabilities.

*Rating update:* Given the clients' reports, the platform updates the servers' ratings according to the *rating update rule*, which is defined as a mapping  $\tau : \Theta \times \Theta \times \{0, 1\} \times \mathcal{A} \rightarrow \Delta(\Theta)$ . For example,  $\tau(\theta'_s | \theta_c, \theta_s, r, \alpha_0)$  is the probability of the server's updated rating being  $\theta'_s$ , given the client's rating  $\theta_c$ , the server's own rating  $\theta_s$ , the client's report  $r$ , and the recommended plan  $\alpha_0$ . We focus on the following class of rating update rules (see Fig. 2 for illustration):

$$\tau(\theta'_s | \theta_c, \theta_s, r, \alpha_0) = \begin{cases} \beta_{\theta_s}^+, & \theta'_s = 1, r \geq \alpha_0(\theta_c, \theta_s) \\ 1 - \beta_{\theta_s}^+, & \theta'_s = 0, r \geq \alpha_0(\theta_c, \theta_s) \\ 1 - \beta_{\theta_s}^-, & \theta'_s = 1, r < \alpha_0(\theta_c, \theta_s) \\ \beta_{\theta_s}^-, & \theta'_s = 0, r < \alpha_0(\theta_c, \theta_s) \end{cases}.$$

In the above rating update rule, if the reported service quality is not lower than the recommended service quality, a server with rating  $\theta_s$  will have rating 1 with probability  $\beta_{\theta_s}^+$ ; otherwise, it will have rating 0 with probability  $\beta_{\theta_s}^-$ . Other more elaborate rating update rules may be considered. But we show that this simple one is good enough to achieve the social optimum.

*Recommended strategy:* The final key component of the rating mechanism is the recommended *strategy*, which determines what recommended plan should be announced in each period. In each period  $t$ , the mechanism keeps track of the history of rating distributions, denoted by  $\mathbf{h}^t = (\mathbf{s}^0, \dots, \mathbf{s}^t) \in S^{t+1}$ , and chooses the recommended plan based on  $\mathbf{h}^t$ . In other words, the recommended strategy is a mapping from the set of histories to its plan set, denoted by  $\pi_0 : \cup_{t=0}^{\infty} S^{t+1} \rightarrow \mathcal{A}$ . Denote the set of all recommended strategies by  $\Pi$ . Note that although the

TABLE III. GIFT-GIVING GAME BETWEEN A CLIENT AND A SERVER.

	high-quality	low-quality
request	$(b, -c)$	$(0, 0)$

rating mechanism knows the rating profile, it determines the recommended plan based on the history of rating distributions, because 1) this reduces the computational and memory complexity of the protocol, and 2) it is easy for the users to follow since they do not know the rating profile. Moreover, since the plan set  $\mathcal{A}$  has 16 elements, the complexity of choosing the plan is large. Hence, we consider the strategies that choose plans from a subset  $\mathcal{B} \subseteq \mathcal{A}$ , and define  $\Pi(\mathcal{B})$  as the set of strategies restricted on the subset  $\mathcal{B}$  of plans.

In summary, the rating mechanism can be represented by the design parameters: the rating update rule and the recommended strategy, and can be denoted by the tuple  $(\tau, \pi_0)$ .

2) *Payoffs*: Once a server and a client are matched, they play the gift-giving game in Table III [2]–[7][11][13], where the row player is the client and the column player is the server. We normalize the payoffs received by the client and the server when a server provides low-quality services to 0. When a server provides high-quality services, the client gets a benefit of  $b > 0$  and the worker incurs a cost of  $c \in (0, b)$ . In the unique Nash equilibrium of the gift-giving game, the server will provide low-quality services, which results in a zero payoff for both the client and the server. Note that as in [2]–[7][11][13], we assume that the same gift-giving game is played for all the client-server pairs. This assumption is reasonable when the number of users is large. Since  $b$  can be considered as a user’s expected benefit across different servers, and  $c$  as its expected cost of high-quality service across different clients, the users’ expected benefits/costs should be approximately the same when the number of users is large. This assumption is also valid when the users have different realized benefit and cost in each period but the same expected benefit  $b$  and expected cost  $c$  across different periods.

*Expected payoff in one period*: Based on the gift-giving game, we can calculate each user’s expected payoff obtained in one period. A user’s expected payoff in one period depends on its own rating, the rating distribution, and the users’ plans. We write user  $i$ ’s plan as  $\alpha_i \in \mathcal{A}$ , and the plan profile of all the users as  $\alpha = (\alpha_1, \dots, \alpha_N)$ . Then user  $i$ ’s expected payoff in one period is  $u_i(\theta_i, \mathbf{s}, \alpha)$ . For illustration, we calculate the users’ expected payoffs under several important

scenarios, assuming that all the users follow the recommended plan (i.e.  $\alpha_i = \alpha_0, \forall i \in \mathcal{N}$ ). When the altruistic plan  $\alpha^a$  is recommended, all the users receive the same expected payoff in one period as

$$u_i(\theta_i, \mathbf{s}, \alpha^a \cdot \mathbf{1}_N) = b - c, \quad \forall i, \theta_i, \mathbf{s},$$

where  $\alpha \cdot \mathbf{1}_N$  is the plan profile in which every user chooses plan  $\alpha$ . Similarly, when the selfish plan  $\alpha^s$  is recommended, all the users receive zero expected payoff in one period, namely

$$u_i(\theta_i, \mathbf{s}, \alpha^s \cdot \mathbf{1}_N) = 0, \quad \forall i, \theta_i, \mathbf{s}.$$

When the fair plan  $\alpha^f$  is recommended, the users receive expected payoffs in one period as follows

$$u_i(\theta_i, \mathbf{s}, \alpha^f \cdot \mathbf{1}_N) = \begin{cases} \frac{s_0-1}{N-1} \cdot b - c, & \theta_i = 0 \\ b - \frac{s_1-1}{N-1} \cdot c, & \theta_i = 1 \end{cases}. \quad (5)$$

Under the fair plan, the users with rating 1 receive a payoff higher than  $b - c$ , because they get high-quality services from everyone and provide high-quality services only when matched to clients with rating 1. In contrast, the users with rating 0 receive a payoff lower than  $b - c$ . Hence, the fair plan  $\alpha^f$  can be considered as a differential punishment.

*Discounted average payoff:* Each user  $i$  has its own strategy  $\pi_i \in \Pi$ . Write the joint strategy profile of all the users as  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_N)$ . Then given the initial rating profile  $\boldsymbol{\theta}^0$ , the recommended strategy  $\pi_0$  and the joint strategy profile  $\boldsymbol{\pi}$  induce a probability distribution over the sequence of rating profiles  $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \dots$ . Taking the expectation with respect to this probability distribution, each user  $i$  receives a discounted average payoff  $U_i(\boldsymbol{\theta}^0, \pi_0, \boldsymbol{\pi})$  calculated as

$$U_i(\boldsymbol{\theta}^0, \pi_0, \boldsymbol{\pi}) = \mathbb{E}_{\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \dots} \left\{ (1 - \delta) \sum_{t=0}^{\infty} \delta^t u_i(\theta_i^t, \mathbf{s}(\boldsymbol{\theta}^t), \boldsymbol{\pi}(\mathbf{s}(\boldsymbol{\theta}^0), \dots, \mathbf{s}(\boldsymbol{\theta}^t))) \right\}$$

where  $\delta \in [0, 1)$  is the common discount factor of all the users. The discount factor  $\delta$  is the rate at which the users discount future payoffs, and reflects the patience of the users. A more patient user has a larger discount factor. Note that the recommended strategy  $\pi_0$  does affect the users' discounted average payoffs by affecting the evolution of the rating profile (i.e. by affecting the expectation operator  $\mathbb{E}_{\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \dots}$ ).

3) *Definition of The Equilibrium:* The platform adopts *sustainable rating mechanisms*, which specifies a tuple of rating update rule and recommended strategy  $(\tau, \pi_0)$ , such that the users find it in their self-interests to follow the recommended strategy. In other words, the recommended strategy should be an equilibrium.

Note that the interaction among the users is neither a repeated game [17] nor a standard stochastic game [18]. In a repeated game, every stage game is the same, which is clearly not true in the platform because users' stage-game payoff  $u_i(\theta_i, \mathbf{s}, \boldsymbol{\alpha})$  depends on the rating distribution  $\mathbf{s}$ . In a standard stochastic game, the state must satisfy: 1) the state and the plan profile uniquely determines the stage-game payoff, and 2) the state is known to all the users. In the platform, the user's stage-game payoff  $u_i(\theta_i, \mathbf{s}, \boldsymbol{\alpha})$  depends on its own rating  $\theta_i$ , which should be included in the state and be known to all the users. Hence, if we were to model the interaction as a standard stochastic game, we need to define the state as the rating profile  $\boldsymbol{\theta}$ . However, the rating profile is not known to the users in our formulation.

To reflect our restriction on recommended strategies that depend only on rating distributions, we define the equilibrium as public announcement equilibrium (PAE), since the strategy depends on the publicly announced rating distributions. Before we define PAE, we need to define the continuation strategy  $\pi|_{\mathbf{h}^t}$ , which is a mapping  $\pi|_{\mathbf{h}^t} : \cup_{k=0}^{\infty} \mathcal{H}^k \rightarrow A$  with  $\pi|_{\mathbf{h}^t}(\mathbf{h}^k) = \pi(\mathbf{h}^t \mathbf{h}^k)$ , where  $\mathbf{h}^t \mathbf{h}^k$  is the concatenation of  $\mathbf{h}^t$  and  $\mathbf{h}^k$ .

*Definition 1:* A pair of a recommended strategy and a symmetric strategy profile  $(\pi_0, \pi_0 \cdot \mathbf{1}_N)$  is a PAE, if for all  $t \geq 0$ , for all  $\tilde{\mathbf{h}}^t \in \mathcal{H}^t$ , and for all  $i \in \mathcal{N}$ , we have

$$U_i(\tilde{\boldsymbol{\theta}}^t, \pi_0|_{\tilde{\mathbf{h}}^t}, \pi_0|_{\tilde{\mathbf{h}}^t} \cdot \mathbf{1}_N) \geq U_i(\tilde{\boldsymbol{\theta}}^t, \pi_0|_{\tilde{\mathbf{h}}^t}, (\pi_i|_{\tilde{\mathbf{h}}^t}, \pi_0|_{\tilde{\mathbf{h}}^t} \cdot \mathbf{1}_{N-1})), \quad \forall \pi_i|_{\tilde{\mathbf{h}}^t} \in \Pi,$$

where  $(\pi_i|_{\tilde{\mathbf{h}}^t}, \pi_0|_{\tilde{\mathbf{h}}^t} \cdot \mathbf{1}_{N-1})$  is the continuation strategy profile in which user  $i$  deviates to  $\pi_i|_{\tilde{\mathbf{h}}^t}$  and the other users follow the strategy  $\pi_0|_{\tilde{\mathbf{h}}^t}$ .

Note that in the definition, we allow a user to deviate to any strategy  $\pi_i \in \Pi$ , even if the recommended strategy  $\pi_0$  is restricted to a subset  $\mathcal{B}$  of plans. Hence, the rating mechanism is robust, in the sense that a user cannot gain even when it uses more complicated strategies. Note also that although a rating mechanism can choose the initial rating profile  $\boldsymbol{\theta}^0$ , we require a recommended strategy to fulfill the incentive constraints under all the initial rating profiles. This adds to the flexibility in choosing the initial rating profile.

PAE is stronger than the Nash equilibrium (NE), because PAE requires the users to not deviate

following *any* history, while NE requires the users to not deviate following the histories that happen in the equilibrium. In this sense, PAE can be considered as a special case of public perfect equilibrium (PPE) in standard repeated and stochastic games, where the strategies depend only on the rating distribution.

### B. The Rating Protocol Design Problem

The goal of the rating mechanism designer is to choose a rating mechanism  $(\tau, \pi_0)$ , such that the social welfare at the equilibrium is maximized in the worst case (with respect to different initial rating profiles). Maximizing the worst-case performance gives us a much stronger performance guarantee than maximizing the performance under a given initial rating profile. Given the rating update error  $\varepsilon$ , the discount factor  $\delta$ , and the subset  $\mathcal{B}$  of plans, the rating mechanism design problem is formulated as:

$$W(\varepsilon, \delta, \mathcal{B}) = \max_{\tau, \pi_0 \in \Pi(\mathcal{B})} \min_{\theta^0 \in \Theta^N} \frac{1}{N} \sum_{i \in \mathcal{N}} U_i(\theta^0, \pi_0, \pi_0 \cdot \mathbf{1}_N) \\ \text{s.t.} \quad (\pi_0, \pi_0 \cdot \mathbf{1}_N) \text{ is a PAE.} \quad (6)$$

Note that  $W(\varepsilon, \delta, \mathcal{B})$  is strictly smaller than the social optimum  $b - c$  for any  $\varepsilon$ ,  $\delta$ , and  $\mathcal{B}$ . This is because to exactly achieve  $b - c$ , the protocol must recommend the altruistic plan  $\alpha^a$  all the time (even when someone shirks), which cannot be an equilibrium. However, we can design rating mechanisms such that for any fixed  $\varepsilon \in [0, 0.5)$ ,  $W(\varepsilon, \delta, \mathcal{B})$  can be arbitrarily close to the social optimum. In particular, such rating mechanisms can be simple, in that  $\mathcal{B}$  can be a small subset of three plans (i.e.  $\mathcal{B} = A^{\text{afs}} \triangleq \{\alpha^a, \alpha^f, \alpha^s\}$ ).

## IV. SOURCES OF INEFFICIENCY

To illustrate the importance of designing optimal, yet simple rating schemes, as well as the challenges associated with determining such a design, in this section, we discuss several simple rating mechanisms that appear to work well intuitively, and show that they are actually bounded away from the social optimum even when the users are arbitrarily patient. We will illustrate why they are inefficient, which gives us some insights on how to design socially-optimal rating mechanisms.



### A. Stationary Recommended Strategies

1) *Analysis*: We first consider rating mechanisms with stationary recommended strategies, which determine the recommended plan solely based on the current rating distribution. Since the game is infinitely-repeatedly played, given the same rating distribution, the continuation game is the same regardless of when the rating distribution occurs. Hence, similar to MDP, we can assign value functions  $V_{\theta}^{\pi_0} : S \rightarrow \mathbb{R}$ ,  $\forall \theta$  for a stationary strategy  $\pi_0$ , with  $V_{\theta}^{\pi_0}(\mathbf{s})$  being the continuation payoff of a user with rating  $\theta$  at the rating distribution  $\mathbf{s}$ . Then, we have the following set of equalities that the value function needs to satisfy:

$$\begin{aligned} V_{\theta_i}^{\pi_0}(\mathbf{s}) &= (1 - \delta) \cdot u_i(\pi_0(\mathbf{s}), \pi_0(\mathbf{s}) \cdot \mathbf{1}_N) \\ &+ \delta \cdot \sum_{\theta'_i, \mathbf{s}'} \Pr(\theta'_i, \mathbf{s}' | \theta_i, \mathbf{s}, \pi_0(\mathbf{s}), \pi_0(\mathbf{s}) \cdot \mathbf{1}_N) \cdot V_{\theta'_i}^{\pi_0}(\mathbf{s}'), \forall i \in \mathcal{N}, \end{aligned} \quad (7)$$

where  $\Pr(\theta'_i, \mathbf{s}' | \theta_i, \mathbf{s}, \pi_0(\mathbf{s}), \pi_0(\mathbf{s}) \cdot \mathbf{1}_N)$  is the transition probability. We can solve for the value function from the above set of equalities, which are similar to the Bellman equations in MDP. However, note that obtaining the value function is not the final step. We also need to check the incentive compatibility (IC) constraints. For example, to prevent user  $i$  from deviating to plan  $\alpha'$ , the following inequality has to be satisfied:

$$\begin{aligned} V_{\theta_i}^{\pi_0}(\mathbf{s}) &\geq (1 - \delta) \cdot u_i(\pi_0(\mathbf{s}), (\alpha', \pi_0(\mathbf{s}) \cdot \mathbf{1}_{N-1})) \\ &+ \delta \cdot \sum_{\theta'_i, \mathbf{s}'} \Pr(\theta'_i, \mathbf{s}' | \theta_i, \mathbf{s}, \pi_0(\mathbf{s}), (\alpha', \pi_0(\mathbf{s}) \cdot \mathbf{1}_{N-1})) \cdot V_{\theta'_i}^{\pi_0}(\mathbf{s}'), \forall i \in \mathcal{N}. \end{aligned} \quad (8)$$

Given a rating mechanism with a stationary recommended strategy  $\pi_0$ , if its value function satisfies all the IC constraints, we can determine the social welfare of the rating mechanism. For example, suppose that all the users have an initial rating of 1. Then, all of them achieve the expected payoff  $V_1^{\pi_0}(0, N)$ , which is the social welfare achieved under this rating mechanism.

Note that given a recommended strategy  $\pi_0$ , it is not difficult to compute the value function by solving the set of linear equations in (7) and check the IC constraints according to the set of linear inequalities in (8). However, it is difficult to derive structural results on the value function (e.g. whether the state with more rating-1 users has a higher value), and thus difficult to know the structures of the optimal recommended strategy (e.g. whether the optimal recommended strategy is a threshold strategy). The difficulty mainly comes from the complexity of the transition

probabilities  $\Pr(\theta'_i, \mathbf{s}' | \theta_i, \mathbf{s}, \pi_0(\mathbf{s}), \pi_0(\mathbf{s}) \cdot \mathbf{1}_N)$ . For example, assuming  $\pi_0(\mathbf{s}) = \alpha^{\mathbf{a}}$ , we have

$$\Pr(1, \mathbf{s}' | 1, \alpha^{\mathbf{a}}, \alpha^{\mathbf{a}} \cdot \mathbf{1}_N) = x_1^+.$$

$$\sum_{k=\max\{0, s'_1-1-(N-s_1)\}}^{\min\{s_1-1, s'_1-1\}} \binom{s_1-1}{k} (x_1^+)^k (1-x_1^+)^{s_1-1-k} \binom{N-s_1}{s'_1-1-k} (x_0^+)^{s'_1-1-k} (1-x_0^+)^{N-s_1-(s'_1-1-k)},$$

where  $x_1^+ \triangleq (1-\varepsilon)\beta_1^+ + \varepsilon(1-\beta_1^-)$  is the probability that a rating-1 user's rating remains to be 1, and  $x_0^+ \triangleq (1-\varepsilon)\beta_0^+ + \varepsilon(1-\beta_0^-)$  is the probability that a rating-0 user's rating goes up to 1. We can see that the transition probability has combinatorial numbers in it and is complicated. Hence, although the stationary strategies themselves are simpler than the nonstationary strategies, they are *harder to compute*, in the sense that it is difficult to derive structural results for rating mechanisms with stationary recommended strategy. In contrast, we are able to develop a unified design framework for socially-optimal rating mechanisms with nonstationary recommended strategies.

2) *Inefficiency*: We measure the efficiency of the rating mechanisms with stationary recommended strategies using the ‘‘price of stationarity’’ (PoStat), defined as

$$\text{PoStat}(\varepsilon, \mathcal{B}) = \frac{\lim_{\delta \rightarrow 1} W^s(\varepsilon, \delta, \mathcal{B})}{b-c}, \quad (9)$$

where  $W^s(\varepsilon, \delta, \mathcal{B})$  is the optimal value of a modified optimization problem (6) with an additional constraint that  $\pi_0$  is stationary.

Note that  $\text{PoStat}(\varepsilon, \mathcal{B})$  measures the efficiency of a class of rating mechanisms (not a specific rating mechanism), because we optimize over all the rating update rules and stationary recommended strategies restricted on  $\mathcal{B}$ . PoStat is a number between 0 and 1. A small PoStat indicates a low efficiency.

Through simulation, we can compute  $\text{PoStat}(0.1, A^{\text{afs}}) = 0.720$ . In other words, even with differential punishment  $\alpha^{\text{f}}$ , the performance of stationary strategies is bounded away from social optimum. We compute the PoStat in a platform with  $N = 5$  users, the benefit  $b = 3$ , the cost  $c = 1$ , and  $\varepsilon = 0.1$ . Under each discount factor  $\delta$ , we assign values between 0 and 1 with a 0.1 grid to  $\beta_\theta^+, \beta_\theta^-$  in the rating update rule, namely we try  $11^4$  rating update rules to select the optimal one. For each rating update rule, we try all the  $3^{N+1} = 729$  stationary recommended strategies restricted on  $A^{\text{afs}}$ . In Table IV, we list normalized social welfare under different discount factors.

As mentioned before, the inefficiency of stationary strategies is due to the punishment exerted under certain rating distributions. For example, the optimal recommended strategies discussed

TABLE IV. NORMALIZED SOCIAL WELFARE OF STATIONARY STRATEGIES RESTRICTED ON  $A^{\text{afs}}$ .

$\delta$	0.7	0.8	0.9	0.99	0.999	0.9999
Normalized welfare	0.690	0.700	0.715	0.720	0.720	0.720

TABLE V. MINIMUM PUNISHMENT PROBABILITIES OF RATING MECHANISMS RESTRICTED ON  $A^{\text{afs}}$  WHEN  $\varepsilon = 0.1$ .

$\delta$	0.7	0.8	0.9	0.99	0.999	0.9999	0.99999
Minimum $\beta_1^-$	0.8	0.8	0.6	0.6	0.3	0.3	0.3

above recommend the selfish or fair plan when at least one user has rating 0, resulting in performance loss. One may think that when the users are more patient (i.e. when the discount factor is larger), we can use milder punishments by lowering the punishment probabilities  $\beta_1^-$  and  $\beta_0^-$ , such that the rating distributions with many low-rating users happen with less frequency. However, simulations on the above strategies show that, to fulfill the IC constraints, the punishment probabilities cannot be made arbitrarily small even when the discount factor is large. For example, Table V shows the minimum punishment probability  $\beta_1^-$  (which is smaller than  $\beta_0^-$ ) of rating mechanisms restricted on  $A^{\text{afs}}$  under different discount factors. In other words, the rating distributions with many low-rating users will happen with some probabilities bounded above zero, with a bound independent of the discount factor. Hence, the performance loss is bounded above zero regardless of the discount factor. Note that in a nonstationary strategy, we could choose whether to punish in rating distributions with many low-rating users, depending on the history of past rating distributions. This adaptive adjustment of punishments allows nonstationary strategies to potentially achieve the social optimum.

### B. Lack of Differential Punishments

We have discussed in the previous subsection the inefficiency of stationary strategies. Now we consider a class of nonstationary strategies restricted on the subset of plans  $A^{\text{as}}$ . Under this class of strategies, all the users are rewarded (by choosing  $\alpha^{\text{a}}$ ) or punished (by choosing  $\alpha^{\text{s}}$ ) simultaneously. In other words, there is no differential punishment that can “transfer” some payoff from low-rating users to high-rating users. We quantify the performance loss of this class of nonstationary strategies restricted on  $A^{\text{as}}$  as follows.

*Proposition 1:* For any  $\varepsilon > 0$ , we have

$$\lim_{\delta \rightarrow 1} W(\varepsilon, \delta, A^{\text{as}}) \leq b - c - \zeta(\varepsilon), \quad (10)$$

where  $\zeta(\varepsilon) > 0$  for any  $\varepsilon > 0$ .

*Proof:* The proof is similar to the proof of [6, Proposition 6]; see Appendix A. ■

The above proposition shows that the maximum social welfare achievable by  $(\pi_0, \pi \cdot \mathbf{1}_N) \in \Pi(A^{\text{as}}) \times \Pi^N(A^{\text{as}})$  at the equilibrium is bounded away from the social optimum  $b - c$ , unless there is no rating update error. Note that the performance loss is independent of the discount factor. In contrast, we will show later that, if we can use the fair plan  $\alpha^f$ , the social optimum can be asymptotically achieved when the discount factor goes to 1. Hence, the differential punishment introduced by the fair plan is crucial for achieving the social optimum.

## V. SOCIALLY OPTIMAL DESIGN

In this section, we design rating mechanisms that asymptotically achieve the social optimum at the equilibrium, even when the rating update rule  $\varepsilon > 0$ . In our design, we use the APS technique, named after the authors of the seminal paper [16], which is also used to prove the folk theorem for repeated games in [17] and for stochastic games in [18]. We will briefly introduce the APS technique first. Meanwhile, more importantly, we will illustrate why we *cannot* use APS in our setting in the same way as [17] and [18] did. Then, we will show how we use APS in a different way in our setting, in order to design the optimal rating mechanism and to construct the equilibrium strategy. Finally, we analyze the performance of a class of simple but suboptimal strategies, which sheds light on why the proposed strategy can achieve the social optimum.

### A. The APS Technique

APS [16] provides a characterization of the set of PPE payoffs. It builds on the idea of *self-generating sets* described as follows. Note that APS is used for standard stochastic games, and recall from our discussion in Section II that the state of the standard stochastic game is the rating profile  $\theta$ . Then define a set  $\mathcal{W}^\theta \subset \mathbb{R}^N$  for every state  $\theta \in \Theta^N$ , and write  $(\mathcal{W}^{\theta'})_{\theta' \in \Theta^N}$  as the collection of these sets. Then we have the following definitions [16][18][19]. First, we say a payoff profile  $\mathbf{v}(\theta) \in \mathbb{R}^N$  is *decomposable* on  $(\mathcal{W}^{\theta'})_{\theta' \in \Theta^N}$  given  $\theta$ , if there exists a recommended

plan  $\alpha_0$ , an plan profile  $\alpha^*$ , and a *continuation payoff function*  $\gamma : \Theta^N \rightarrow \cup_{\theta' \in \Theta^N} \mathcal{W}^{\theta'}$  with  $\gamma(\theta') \in \mathcal{W}^{\theta'}$ , such that for all  $i \in \mathcal{N}$  and for all  $\alpha_i \in A$ ,

$$\begin{aligned} v_i &= (1 - \delta)u_i(\boldsymbol{\theta}, \alpha_0, \alpha^*) + \delta \sum_{\theta'} \gamma_i(\theta')q(\theta'|\boldsymbol{\theta}, \alpha_0, \alpha^*) \\ &\geq (1 - \delta)u_i(\boldsymbol{\theta}, \alpha_0, \alpha_i, \alpha_{-i}^*) + \delta \sum_{\theta'} \gamma_i(\theta')q(\theta'|\boldsymbol{\theta}, \alpha_0, \alpha_i, \alpha_{-i}^*). \end{aligned} \quad (11)$$

Then, we say a set  $(\mathcal{W}^{\theta})_{\theta \in \Theta^N}$  is a self-generating set, if for any  $\boldsymbol{\theta}$ , every payoff profile  $\mathbf{v}(\boldsymbol{\theta}) \in \mathcal{W}^{\boldsymbol{\theta}}$  is decomposable on  $(\mathcal{W}^{\theta'})_{\theta' \in \Theta^N}$  given  $\boldsymbol{\theta}$ . The important property of self-generating sets is that any self-generating set is a set of PPE payoffs [16][18][19].

Based on the idea of self-generating sets, [17] and [18] proved the folk theorem for repeated games and stochastic games, respectively. However, we cannot use APS in the same way as [17] and [18] did for the following reason. We assume that the users do not know the rating profile of every user, and restrict our attention on symmetric PA strategy profiles. This requires that each user  $i$  cannot use the continuation payoff function  $\gamma_i(\boldsymbol{\theta})$  directly. Instead, each user  $i$  should assign the same continuation payoff for the rating profiles that have the same rating distribution, namely  $\gamma_i(\boldsymbol{\theta}) = \gamma_i(\boldsymbol{\theta}')$  for all  $\boldsymbol{\theta}$  and  $\boldsymbol{\theta}'$  such that  $\mathbf{s}(\boldsymbol{\theta}) = \mathbf{s}(\boldsymbol{\theta}')$ .

### B. Socially Optimal Design

As mentioned before, the social optimum  $b - c$  can be exactly achieved only by servers providing high-quality service all the time, which is not an equilibrium. Hence, we aim at achieving the social optimum  $b - c$  asymptotically. We define the asymptotically socially optimal rating mechanisms as follows.

*Definition 2 (Asymptotically Socially Optimal Rating Mechanisms):* Given a rating update error  $\varepsilon \in [0, 0.5)$ , we say a rating mechanism  $(\tau(\varepsilon), \pi_0(\varepsilon, \xi, \delta) \in \Pi)$  is asymptotically socially optimal under  $\varepsilon$ , if for any small performance loss  $\xi > 0$ , we can find a  $\underline{\delta}(\xi)$ , such that for any discount factor  $\delta > \underline{\delta}(\xi)$ , we have

- $(\pi_0(\xi, \delta), \pi_0(\xi, \delta) \cdot \mathbf{1}_N)$  is a PAE;
- $U_i(\boldsymbol{\theta}^0, \pi_0, \pi_0 \cdot \mathbf{1}_N) \geq b - c - \xi, \forall i \in \mathcal{N}, \forall \boldsymbol{\theta}^0$ .

Note that in the asymptotically socially optimal rating mechanism, the rating update rule depends only on the rating update error, and works for any tolerated performance loss  $\xi$  and for any the discount factor  $\delta > \underline{\delta}$ . The recommended strategy  $\pi_0$  is a class of strategies parameterized

by  $(\varepsilon, \xi, \delta)$ , and works for any  $\varepsilon \in [0, 0.5)$ , any  $\xi > 0$  and any discount factor  $\delta > \underline{\delta}$  under the rating update rule  $\tau(\varepsilon)$ .

First, we define a few auxiliary variables first for better exposition of the theorem. Define  $\kappa_1 \triangleq \frac{b}{\frac{N-2}{N-1}b-c} - 1$  and  $\kappa_2 \triangleq 1 + \frac{c}{(N-1)b}$ . In addition, we write the probability that a user with rating 1 has its rating remain at 1 if it follows the recommended altruistic plan  $\alpha^a$  as:

$$x_1^+ \triangleq (1 - \varepsilon)\beta_1^+ + \varepsilon(1 - \beta_1^-).$$

Write the probability that a user with rating 1 has its rating remain at 1 if it follows the recommended fair plan  $\alpha^f$  as:

$$x_{s_1(\boldsymbol{\theta})} \triangleq \left[ (1 - \varepsilon) \frac{s_1(\boldsymbol{\theta}) - 1}{N - 1} + \frac{N - s_1(\boldsymbol{\theta})}{N - 1} \right] \beta_1^+ + \left( \varepsilon \frac{s_1(\boldsymbol{\theta}) - 1}{N - 1} \right) (1 - \beta_1^-).$$

Write the probability that a user with rating 0 has its rating increase to 1 if it follows the recommended plan  $\alpha^a$  or  $\alpha^f$ :

$$x_0^+ \triangleq (1 - \varepsilon)\beta_0^+ + \varepsilon(1 - \beta_0^-).$$

*Theorem 1:* Given any rating update error  $\varepsilon \in [0, 0.5)$ ,

- (*Design rating update rules*): A rating update rule  $\tau(\varepsilon)$  that satisfies
  - Condition 1 (following the recommended plan leads to a higher rating):

$$\beta_1^+ > 1 - \beta_1^- \text{ and } \beta_0^+ > 1 - \beta_0^-,$$

- Condition 2 (Enough “reward” for users with rating 1):

$$x_1^+ = (1 - \varepsilon)\beta_1^+ + \varepsilon(1 - \beta_1^-) > \frac{1}{1 + \frac{c}{(N-1)b}},$$

- Condition 3 (Enough “punishment” for users with rating 0):

$$x_0^+ = (1 - \varepsilon)\beta_0^+ + \varepsilon(1 - \beta_0^-) < \frac{1 - \beta_1^+}{\frac{c}{(N-1)b}},$$

can be the rating update rule in a asymptotically socially-optimal rating mechanism.

- (*Optimal recommended strategies*): Given the rating update rule  $\tau(\varepsilon)$  that satisfies the above conditions, any small performance loss  $\xi > 0$ , and any discount factor  $\delta \geq \underline{\delta}(\varepsilon, \xi)$  with  $\underline{\delta}(\varepsilon, \xi)$  defined in Appendix B, the recommended strategy  $\pi_0(\varepsilon, \xi, \delta) \in \Pi(A^{\text{afs}})$  constructed by Table VI is the recommended strategy in a asymptotically socially-optimal rating mechanism.

*Proof:* See Appendix C for the entire proof. We provide a proof sketch here.

The proof builds on the theory of self-generating sets [16], which can be considered as the extension of Bellman equations in dynamic programming to the cases with multiple self-interested users using nonstationary strategies. We can decompose each user  $i$ 's discounted average payoff into the current payoff and the continuation payoff as follows:

$$\begin{aligned}
& U_i(\boldsymbol{\theta}^0, \pi_0, \boldsymbol{\pi}) \\
&= \mathbb{E}_{\boldsymbol{\theta}^1, \dots} \left\{ (1 - \delta) \sum_{t=0}^{\infty} \delta^t u_i(\theta_i^t, \mathbf{s}(\boldsymbol{\theta}^t), \boldsymbol{\pi}(\mathbf{s}(\boldsymbol{\theta}^0), \dots, \mathbf{s}(\boldsymbol{\theta}^t))) \right\} \\
&= \underbrace{(1 - \delta) \cdot u_i(\theta_i^0, \mathbf{s}(\boldsymbol{\theta}^0), \boldsymbol{\pi}(\mathbf{s}(\boldsymbol{\theta}^0)))}_{\text{current payoff at } t=0} + \underbrace{\delta \cdot \mathbb{E}_{\boldsymbol{\theta}^2, \dots} \left\{ (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} u_i(\theta_i^t, \mathbf{s}(\boldsymbol{\theta}^t), \boldsymbol{\pi}(\mathbf{s}(\boldsymbol{\theta}^0), \dots, \mathbf{s}(\boldsymbol{\theta}^t))) \right\}}_{\text{continuation payoff starting from } t=1}.
\end{aligned}$$

We can see that the continuation payoff starting from  $t = 1$  is the discounted average payoff as if the system starts from  $t = 1$ . Suppose that the users follow the recommended strategy. Since the recommended strategy and the rating update rule do not differentiate users with the same rating, we can prove that the users with the same rating have the same continuation payoff starting from any point. Hence, given  $\pi_0$  and  $\boldsymbol{\pi} = \pi_0 \cdot \mathbf{1}_N$ , the decomposition above can be simplified into

$$v_{\theta}^{\pi_0}(\mathbf{s}) = (1 - \delta) \cdot u(\theta, \mathbf{s}, \alpha_0 \cdot \mathbf{1}_N) + \delta \cdot \sum_{\theta'=0}^1 \sum_{\mathbf{s}'} q(\theta', \mathbf{s}' | \theta, \mathbf{s}, \alpha_0 \cdot \mathbf{1}_N) \cdot v_{\theta'}^{\pi_0}(\mathbf{s}'), \quad (12)$$

where  $q(\theta', \mathbf{s}' | \theta, \mathbf{s}, \alpha \cdot \mathbf{1}_N)$  is the probability that the user has rating  $\theta'$  and the rating distribution is  $\mathbf{s}'$  in the next period given the user's current rating  $\theta$ , the current rating distribution  $\mathbf{s}$ , and the action profile  $\alpha \cdot \mathbf{1}_N$ , and  $v_{\theta}^{\pi_0}(\mathbf{s})$  is the continuation payoff of the users with rating  $\theta$  starting from the initial rating distribution  $\mathbf{s}$ .

The differences between (12) and the Bellman equations are 1) the ‘‘value’’ of state  $\mathbf{s}$  in (12) is a vector comprised of rating-1 and rating-0 users' values, compared to scalar values in Bellman equations, and 2) the value of state  $\mathbf{s}$  is not fixed in (12), because the action  $\alpha_0$  taken under state  $\mathbf{s}$  is not fixed in a nonstationary strategy (this is also the key difference from the analysis of stationary strategies; see (7) where the action taken at state  $\mathbf{s}$  is fixed to be  $\pi_0(\mathbf{s})$ ). In addition, for an equilibrium recommended strategy, the decomposition needs to satisfy the

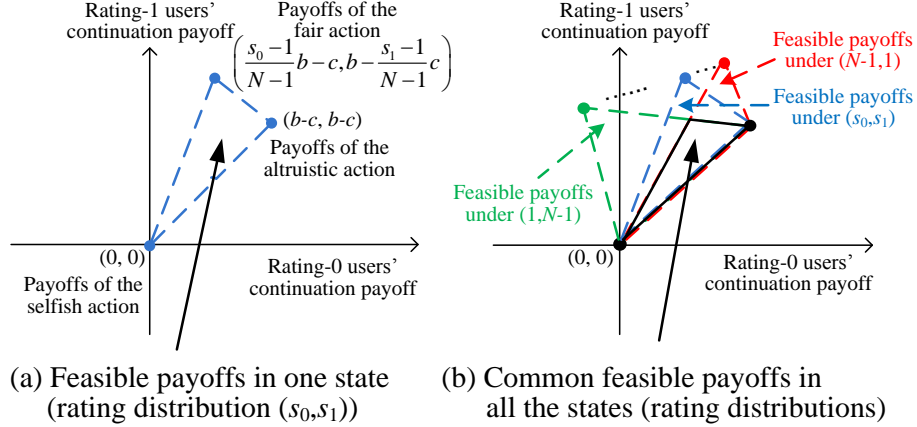


Fig. 3. Illustration of how to build the self-generating set. The left figure shows the set of feasible payoffs in one state (i.e. under the rating distribution  $(s_0, s_1)$ ). The right figure shows the sets of feasible payoffs in different states (i.e. rating distributions) and their intersection, namely the set of common feasible payoffs in all the states (i.e. under all the rating distributions).

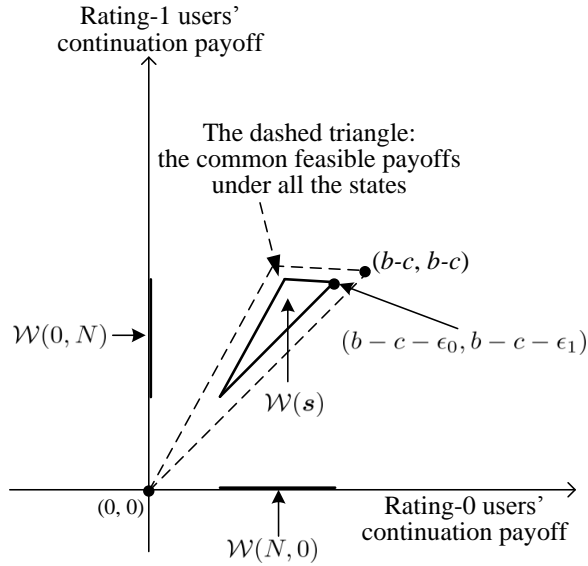


Fig. 4. Illustration of the self-generating set, which is a triangle within the set of common feasible payoffs in Fig. 3.

following incentive constraints: for all  $\alpha$ ,

$$v_{\theta}^{\pi_0}(\mathbf{s}) \geq (1 - \delta) \cdot u(\theta, \mathbf{s}, (\alpha, \alpha_0 \cdot \mathbf{1}_{N-1})) + \delta \cdot \sum_{\theta'=0}^1 \sum_{\mathbf{s}'} \rho(\theta', \mathbf{s}' | \theta, \mathbf{s}, (\alpha, \alpha_0 \cdot \mathbf{1}_{N-1})) \cdot v_{\theta'}^{\pi_0}(\mathbf{s}'). \quad (13)$$

To analyze nonstationary strategies, we use the theory of self-generating sets. Note, however, that [16] does not tell us how to construct a self-generating set, which is exactly the major



difficulty to overcome in our proof. In our proof, we construct the following self-generating set. First, since the strategies depend on rating distributions only, we let  $\mathcal{W}(\boldsymbol{\theta}) = \mathcal{W}(\boldsymbol{\theta}')$  for any  $\boldsymbol{\theta}$  and  $\boldsymbol{\theta}'$  that have the same rating distribution. Hence, in the rest of the proof sketch, we write the self-generating set as  $\{\mathcal{W}(\mathbf{s})\}_{\mathbf{s}}$ , which is illustrated in Fig. 3 and Fig. 4. Fig. 3 shows how to construct the self-generating set. The left of Fig. 3 shows the feasible payoffs in one state, and the right shows the common feasible payoffs in all the states (we consider the common feasible payoffs such that we can use the same  $\mathcal{W}(\mathbf{s})$  under all the states  $\mathbf{s}$ ). The self-generating set is a subset of the common feasible payoffs, as illustrated in Fig. 4. When the users have different ratings (i.e.  $1 \leq s_0 \leq N - 1$ ), the set  $\mathcal{W}(\mathbf{s})$  is the triangle shown in Fig. 4, which is congruent to the triangle of common feasible payoffs (shown in dashed lines), and has the upper right vertex at  $(b - c - \epsilon_0, b - c - \epsilon_1)$  with  $\epsilon_0, \epsilon_1 \leq \xi$ . We have the analytical expression for the triangle in Appendix C. When all the users have the same rating (i.e.  $s_0 = 0$  or  $s_0 = N$ ), only one component in  $\mathbf{v}(\mathbf{s})$  is relevant. Hence, the sets  $\mathcal{W}((N, 0))$  and  $\mathcal{W}((0, N))$  are line segments determined by the ranges of  $v_0$  and  $v_1$  in the triangle, respectively.

In addition, we simplify the decomposition (12) and (13) by letting the continuation payoffs  $v_{\theta'}^{\pi_0}(\mathbf{s}') = v_{\theta'}^{\pi_0}$  for all  $\mathbf{s}'$ . Hence, for a given  $\mathbf{s}$  and a payoff vector  $\mathbf{v}(\mathbf{s})$ , the continuation payoffs  $\mathbf{v}' = (v'_0, v'_1)$  can be determined by solving the following two linear equations:

$$\begin{cases} v_0(\mathbf{s}) = (1 - \delta)u(0, \mathbf{s}, \alpha_0 \mathbf{1}_N) + \delta \sum_{\theta'=0}^1 q(\theta'|0, \alpha_0 \mathbf{1}_N) v'_{\theta'} \\ v_1(\mathbf{s}) = (1 - \delta)u(1, \mathbf{s}, \alpha_0 \mathbf{1}_N) + \delta \sum_{\theta'=0}^1 q(\theta'|1, \alpha_0 \mathbf{1}_N) v'_{\theta'} \end{cases} \quad (14)$$

where  $q(\theta'|\theta, \alpha_0)$  is the probability that the next rating is  $\theta'$  for a user with rating  $\theta$  under the plan profile  $\alpha_0 \cdot \mathbf{1}_N$ .

Based on the above simplification, the collection of sets  $\{\mathcal{W}(\mathbf{s})\}_{\mathbf{s}}$  in Fig. 4 is a self-generating set, if for any  $\mathbf{s}$  and any payoff vector  $\mathbf{v}(\mathbf{s}) \in \mathcal{W}(\mathbf{s})$ , we can find a plan  $\alpha_0$  such that the continuation payoffs  $\mathbf{v}'$  calculated from (14) lie in the triangle and satisfy the incentive constraints in (13).

In summary, we can prove that the collection of sets  $\{\mathcal{W}(\mathbf{s})\}_{\mathbf{s}}$  illustrated in Fig. 4 is a self-generating set under certain conditions. Specifically, given a performance loss  $\xi$ , we construct the corresponding  $\{\mathcal{W}(\mathbf{s})\}_{\mathbf{s}}$ , and prove that it is a self-generating set under the following conditions: 1) the discount factor  $\delta \geq \underline{\delta}(\varepsilon, \xi)$  with  $\underline{\delta}(\varepsilon, \xi)$  defined in Appendix B, and 2) the three conditions on the rating update rule in Theorem 1. This proves the first part of Theorem 1.

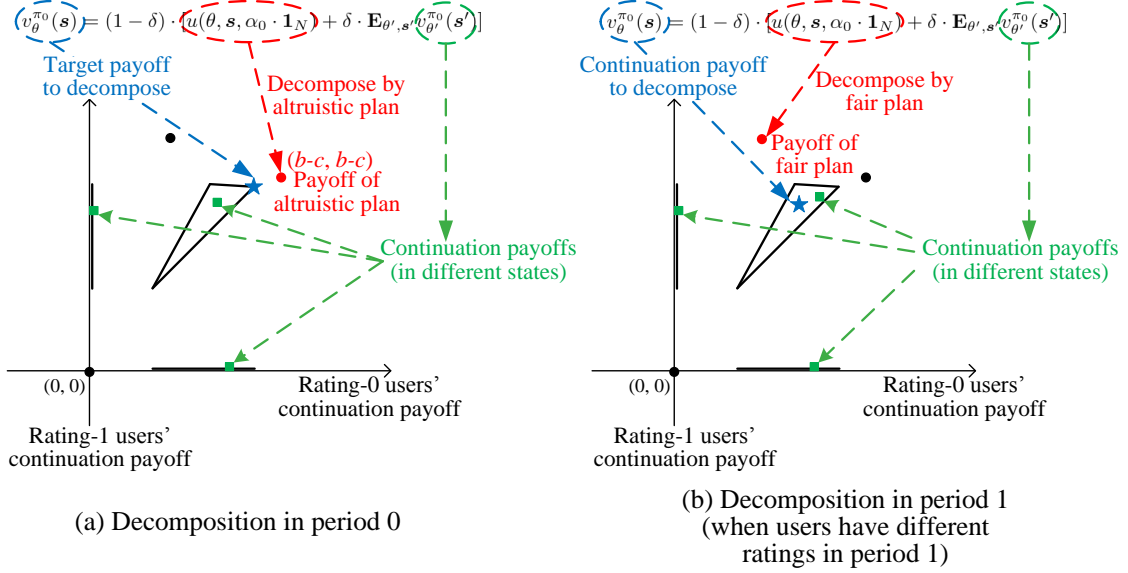


Fig. 5. The decomposition of payoffs. The left figure shows the decomposition in period 0, when the payoff to decompose is the target payoff  $(b - c - \epsilon_0, b - c - \epsilon_1)$ ; the right figure shows the decomposition in period 1, when the payoff to decompose is the continuation payoff starting from period 1 and when the users have different ratings.

The corresponding recommended strategy can be determined based on the decomposition of payoffs. Specifically, given the current rating distribution  $s$  and the current expected payoffs  $v(s) \in \mathcal{W}(s)$ , we find a plan  $\alpha_0$  such that the continuation payoffs  $v'$  calculated from (14) lie in the triangle and satisfy the incentive constraints. The decomposition is illustrated in Fig. 5. One important issue in the decomposition is which plan should be used to decompose the payoff. We prove that we can determine the plan in the following way (illustrated in Fig. 6). When the users have different ratings, choose the altruistic plan  $\alpha^a$  when  $v(s)$  lies in the region marked by “a” in the triangle in Fig. 6-(b), and choose the fair plan  $\alpha^f$  otherwise. When the users have the same rating 0 or 1, choose the altruistic plan  $\alpha^a$  when  $v_0(s)$  or  $v_1(s)$  lies in the region marked by “a” in the line segment in Fig. 6-(a) or Fig. 6-(c), and choose the selfish plan  $\alpha^s$  otherwise. Note that we can analytically determine the line that separates the two regions in the triangle and the threshold that separates the two regions in the line segment (analytical expressions are omitted due to space limitation; see Appendix D for details). The above decomposition is repeated, and is used to determine the recommended plan in each period based on the current rating distribution  $s$  and the current expected payoffs to achieve  $v(s)$ . The procedure described above is exactly

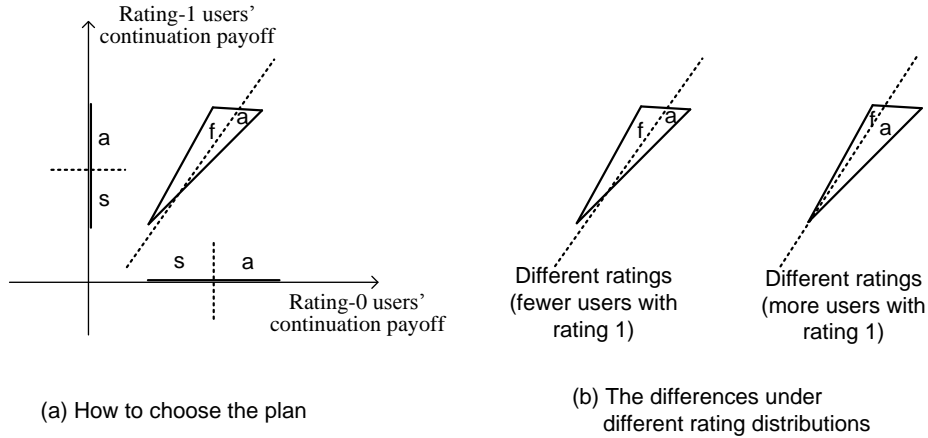


Fig. 6. Illustration of how to choose the plan in order to decompose a given payoff. Each self-generating set is partitioned into two parts. In each period, a recommended plan (the altruistic plan “a”, the fair plan “f”, or the selfish plan “s”) is chosen, depending on which part of the self-generating set the expected payoffs fall into.

the algorithm to construct the recommended strategy, which is described in Table VI. Due to space limitation, Table VI is illustrative but not specific. The detailed table that describes the algorithm can be found in Appendix D. ■

Theorem 1 proves that for any rating update error  $\varepsilon \in [0, 0.5)$ , we can design an asymptotically optimal rating mechanism. The design of the asymptotically optimal rating mechanism consists of two parts. The first part is to design the rating update rule. First, we should give incentives for the users to provide high-quality service, by setting  $\beta_\theta^+$ , the probability that the rating goes up when the service quality is not lower than the recommended quality, to be larger than  $1 - \beta_\theta^-$ , the probability that the rating goes up when the service quality is lower than the recommended quality. Second, for a user with rating 1, the expected probability that its rating goes up when it complies should be larger than the threshold specified in Condition 2 ( $x_{s_1}^+ > x_1^+$  implies that  $x_{s_1}^+$  is larger than the threshold, too). This gives users incentives to obtain rating 1. Meanwhile, for a user with rating 0, the expected probability that its rating goes up when it complies,  $x_0^+$ , should be smaller than the threshold specified in Condition 3. This provides necessary punishment for a user with rating 0. Note that Conditions 2 and 3 imply that  $x_1^+ > x_0^+$ . In this way, a user will prefer to have rating 1.

The second part is to construct the equilibrium recommended strategy. Theorem 1 proves that for any feasible discount factor  $\delta$  no smaller than the lower-bound discount factor  $\underline{\delta}(\varepsilon, \xi)$

TABLE VI. ALGORITHM TO CONSTRUCT RECOMMENDED STRATEGIES.

---



---

**Require:**  $b, c, \varepsilon, \xi; \tau(\varepsilon), \delta \geq \underline{\delta}(\varepsilon, \xi); \theta^0$

---

**Initialization:**  $t = 0, \varepsilon_0 = \xi, \varepsilon_1 = \varepsilon_0 / (1 + \frac{\kappa_2}{\kappa_1}), v^\theta = b - c - \varepsilon_\theta, \theta = \theta^0.$

---

**repeat**

**if**  $s_1(\theta) = 0$  **then**

**if**  $(v^0, v^1)$  lies in region “a” of the horizontal line segment in Fig. 6-(a)

choose recommended plan  $\alpha^a$

**else**

choose recommended plan  $\alpha^s$

**end**

**elseif**  $s_1(\theta) = N$  **then**

**if**  $(v^0, v^1)$  lies in region “a” of the vertical line segment in Fig. 6-(c)

choose recommended plan  $\alpha^a$

**else**

choose recommended plan  $\alpha^s$

**end**

**else**

**if**  $(v^0, v^1)$  lies in region “a” of the triangle in Fig. 6-(b)

choose recommended plan  $\alpha^a$

**else**

choose recommended plan  $\alpha^f$

**end**

**end**

determine the continuation payoffs  $(v'_0, v'_1)$  according to (14)

$t \leftarrow t + 1$ , determine the rating profile  $\theta^t$ , set  $\theta \leftarrow \theta^t, (v_0, v_1) \leftarrow (v'_0, v'_1)$

**until**  $\emptyset$

---



---

defined in Appendix B, we can construct the corresponding recommended strategy such that each user can achieve an discounted average payoff of at least  $b - c - \xi$ . Now we show how to construct the recommended strategy. Note that determining the lower-bound discount factor  $\underline{\delta}(\varepsilon, \xi)$  analytically is important for constructing the equilibrium  $(\pi_0, \pi_0 \cdot \mathbf{1}_N)$ , because a feasible discount factor is needed to determine the strategy. In [17] and [18], the lower bound for the discount factor cannot be obtained analytically. Hence, their results are not constructive.

The algorithm in Table VI that constructs the optimal recommended strategy works as follows. In each period, the algorithm updates the continuation payoffs  $(v_0, v_1)$ , and determines the recommended plan based on the current rating distribution and the continuation payoffs. In Fig. 6, we illustrate which plan to recommend based on where the continuation payoffs locate in the self-generating sets. Specifically, each set  $\mathcal{W}(s)$  is partitioned into two parts (the partition

lines can be determined analytically; see Appendix D for the analytical expressions). When all the users have rating 0 (or 1), we recommend the altruistic plan  $\alpha^a$  if the continuation payoff  $v_0$  (or  $v_1$ ) is large, and the selfish plan  $\alpha^s$  otherwise. When the users have different ratings, we recommend the altruistic plan  $\alpha^a$  when  $(v_0, v_1)$  lies in the region marked by “a” in the triangle in Fig. 6-(b), and the fair plan  $\alpha^f$  otherwise. Note that the partition of  $\mathcal{W}(s)$  is different under different rating distributions (e.g., the region in which the altruistic plan is chosen is larger when more users have rating 1). Fig. 6 also illustrates why the strategy is nonstationary: the recommended plan depends on not only the current rating distribution  $s$ , but also which region of  $\mathcal{V}(s)$  the continuation payoffs  $(v_0, v_1)$  lie in.

**Complexity:** Although the design of the optimal recommended strategy is complicated, the implementation is simple. The computational complexity in each period comes from 1) identifying which region the continuation payoffs lie in, which is simple because the regions are divided by a straight line that is analytically determined, and 2) updating the continuation payoffs  $(v_0, v_1)$  by (14), which can be easily done by solving a set of two linear equations with two variables. The memory complexity is also low: because we summarize the history of past states by the continuation payoffs  $(v_0, v_1)$ , the protocol does not need to store all the past states.

### C. Whitewashing-Proofness

An important issue in rating mechanisms is whitewashing, namely users with low ratings can register as a new user to clear its history of bad behaviors. We say a rating mechanism is whitewashing-proof, if the cost of whitewashing (e.g. creating a new account) is higher than the benefit from whitewashing. The benefit from whitewashing is determined by the difference between the current continuation payoff of a low-rating user and the target payoff of a high-rating user. Since this difference is relatively small under the proposed rating mechanism, the proposed mechanism is robust to whitewashing.

*Proposition 2:* Given the performance loss tolerance  $\xi > 0$ , the proposed rating mechanism is whitewashing-proof if the cost of whitewashing is larger than  $\left(1 - \frac{1}{\kappa_1} - \frac{1}{\kappa_2}\right) \cdot \xi$ .

*Proof:* We illustrate the proof using Fig. 7. In Fig. 7, we show the self-generating set again, and point out the target payoff of a rating-1 user and the lowest continuation payoff of a rating-0 user. The difference between these two payoffs is the highest benefit that a rating-0 user can

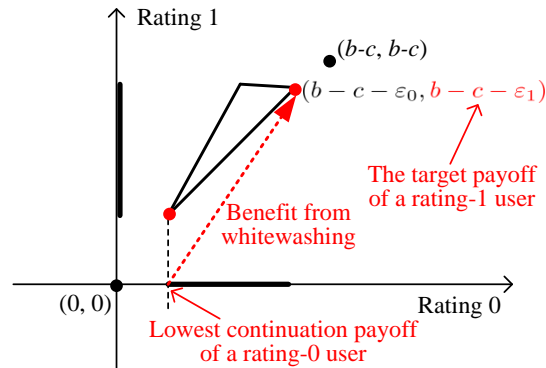


Fig. 7. Illustration of the target payoff of a rating-1 user and the lowest continuation payoff of a rating-0 user.

get by whitewashing. Simple calculation tells us that the difference is  $\left(1 - \frac{1}{\kappa_1} - \frac{1}{\kappa_2}\right) \cdot \xi$ , which completes the proof of Proposition 2. ■

## VI. SIMULATION RESULTS

We compare against the rating mechanism with threshold-based stationary recommended strategies. In particular, we focus on threshold-based stationary recommended strategies that use two plans. In other words, one plan is recommended when the number of rating-1 users is no smaller than the threshold, and the other plan is recommended otherwise. In particular, we consider threshold-based stationary recommended strategies restricted on  $A^{\text{af}}$ ,  $A^{\text{as}}$ , and  $A^{\text{fs}}$ , and call them “Threshold AF”, “Threshold AS”, and “Threshold FS”, respectively. We focus on threshold-based strategies because it is difficult to find the optimal stationary strategy in general when the number of users is large (the number of stationary strategies grows exponentially with the number of users). In our experiments, we fix the following system parameters:  $N = 10, b = 3, c = 1$ .

In Fig. 8, we first illustrate the evolution of the states and the recommended plans taken under the proposed rating mechanism and the rating mechanism with the Threshold AF strategy. The threshold is set to be 5. Hence, it recommends the altruistic plan when at least half of the users have rating 1, and recommends the fair plan otherwise. We can see that in the proposed strategy, the plans taken can be different at the same state. In particular, in “bad” states (6,4) and (7,3) at time slot 3 and 5, respectively, the proposed rating mechanism may recommend the fair plan (as a punishment) and the altruistic plan (i.e. do not punish because the punishment happens

Period	0	1	2	3	4	5	6
Rating distribution	(0,10)	(1,9)	(3,7)	(6,4)	(5,5)	(6,4)	(2,8)
Threshold AF	A	A	A	F	A	F	A
Threshold AS	A	A	A	S	A	S	A
Threshold FS	F	F	F	S	F	S	F
Proposed	A	A	A	F	A	A	A

↑  
The proposed strategy does not punish,  
because users have been punished in period 3

Fig. 8. Evolution of states and recommended plans taken in different rating mechanisms.

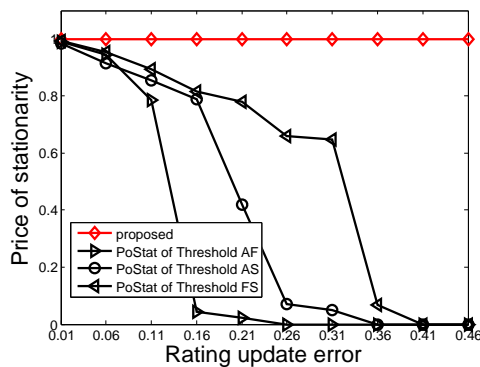


Fig. 9. Price of stationarity of different stationary rating mechanisms under different rating update errors.

in time slot 3), while the stationary mechanism always recommends the fair plan to punish the low-rating users.

Then in Fig. 9, we show the price of stationarity of three representative stationary rating mechanisms: the one with the optimal Threshold AF strategy, the one with the optimal Threshold AS strategy, and the one with the optimal Threshold FS strategy. We can see from Fig. 9 that as the rating update error increases, the efficiency of stationary rating mechanisms decreases, and drops to 0 when the error probability is large (e.g. when  $\varepsilon > 0.4$ ). In contrast, the proposed rating mechanism can achieve arbitrarily close to the social optimum.

In Fig. 10, we illustrate the lower-bound discount factors under different performance loss tolerances and rating update errors. As expected, when the performance loss tolerance becomes larger, the lower-bound discount factor becomes smaller. What is unexpected is how the lower-

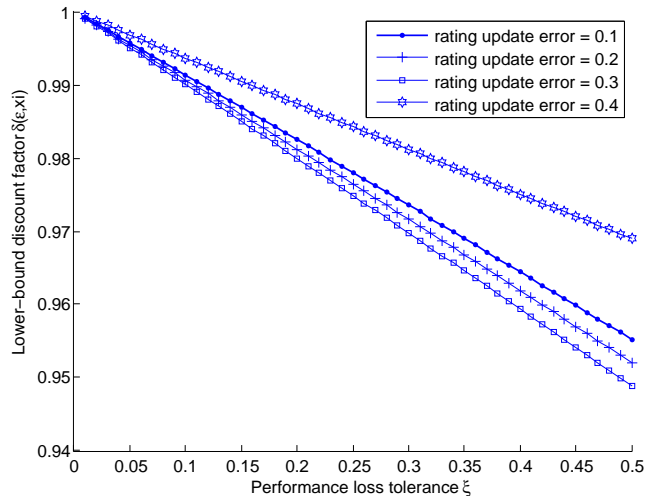


Fig. 10. Lower-bound discount factors under different performance loss tolerances and rating update errors.

bound discount factor changes with the rating update error. Specifically, the lower-bound discount factor decreases initially with the increase of the error, and then increases with the error. It is intuitive to see the discount factor increases with the rating update error, because the users need to be more patient when the rating update is more erroneous. The initial decrease of the discount factor in the error can be explained as follows. If the rating update error is extremely small, the punishment for the rating-0 users in the optimal rating update rule needs to be very severe (i.e. a smaller  $\beta_0^+$  and a larger  $\beta_0^-$ ). Hence, once a user is assigned with rating 0, it needs to be more patient to carry out the severe punishment (i.e. weigh the future payoffs more).

Finally, we illustrate the robustness of the proposed mechanisms with respect to the estimation of rating update errors. Suppose that the rating update error is  $\varepsilon$ . However, the designer cannot accurately measure this error. Under the estimated error  $\hat{\varepsilon}$ , the rating mechanism will construct another recommended strategy. In Fig. 11, we illustrate the performance gain/loss in terms of social welfare under the estimated error  $\hat{\varepsilon}$ , when the rating update error is  $\varepsilon$ . We can see that there is less than 5% performance variance when the estimation inaccuracy is less than 50%. The performance variance is larger when the rating update error is larger.



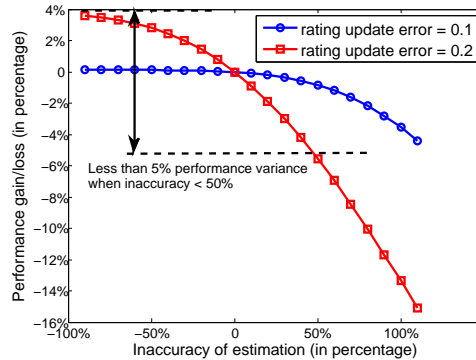


Fig. 11. Performance gain/loss (in percentage) under different inaccuracy of estimation (in percentage).

## VII. CONCLUSION

In this paper, we proposed a design framework for simple binary rating mechanisms that can achieve the social optimum in the presence of rating update errors. We provided design guidelines for the optimal rating update rules, and an algorithm to construct the optimal nonstationary recommended strategy. The key design principles that enable the rating mechanism to achieve the social optimum are the differential punishments, and the nonstationary strategies that reduce the performance loss while providing enough incentives. We also reduced the complexity of computing the recommended strategy by proving that using three recommended plans is enough to achieve the social optimum. The proposed rating mechanism is the first one that can achieve the social optimum even when the rating update errors are large. Simulation results demonstrated the significant performance gain of the proposed rating mechanism over the state-of-the-art mechanisms, especially when the rating update error is large.

## REFERENCES

- [1] A. Kittur, E. Chi, and B. Suh, “Crowdsourcing user studies with mechanical turk,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2008.
- [2] Y. Zhang and M. van der Schaar, “Rating protocols for online communities,” *ACM Transactions on Economics and Computation*, 2013.
- [3] Y. Zhang and M. van der Schaar, “Peer-to-peer multimedia sharing based on social norms,” *Elsevier Journal Signal Processing: Image Communication Special Issue on “Advances in video streaming for P2P networks”*, vol. 27, no. 5, pp. 383–400, May 2012.
- [4] A. Blanc, Y.-K. Liu, and A. Vahdat, “Designing incentives for peer-to-peer routing,” in *Proceedings of the IEEE INFOCOM 2005*, 2005.

- [5] M. Feldman, K. Lai, I. Stoica, and J. Chuang, “Robust incentive techniques for peer-to-peer networks,” in *Proceedings of ACM Conference on Electronic Commerce*, 2004.
- [6] C. Dellarocas, “Reputation mechanism design in online trading environments with pure moral hazard,” *Information Systems Research*, vol. 16, no. 2, pp. 209–230, 2005.
- [7] M. Kandori, “Social norms and community enforcement,” *Review of Economic Studies*, vol. 59, no. 1, pp. 63 – 80, 1992.
- [8] M. Okuno-Fujiwara and A. Postlewaite, “Social norms and random matching games,” *Games and Economic Behaviors*, vol. 9, no. 1, pp. 79 – 109, 1993.
- [9] P. Dal Bó, “Social norms, cooperation and inequality,” *Economic Theory*, vol. 30, pp. 89 – 105, 2007.
- [10] K. Hasker, “Social norms and choice: A general folk theorem for repeated matching games with endogenous matching rules,” *International Journal of Game Theory*, vol. 36, pp. 137–146, 2007.
- [11] S. Takahashi, “Community enforcement when players observe partners’ past play,” *Journal of Economic Theory*, vol. 145, no. 1, pp. 42–62, 2010.
- [12] J. Deb, “Cooperation and community responsibility: A folk theorem for random matching games with names,” *Revised and Resubmitted to the Review of Economic Studies*, 2013.
- [13] G. Ellison, “Cooperation in the prisoner’s dilemma with anonymous random matching,” *Review of Economic Studies*, vol. 61, no. 3, pp. 567 – 588, 1994.
- [14] R. Izhak-Ratzin, H. Park, and M. van der Schaar, “Reinforcement learning in BitTorrent systems,” in *Proc. IEEE Infocom 2011*, 2011.
- [15] H. Park and M. van der Schaar, “Evolution of resource reciprocation strategies in P2P networks,” *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1205–1218, Mar. 2010.
- [16] D. Abreu, D. Pearce, and E. Stacchetti, “Toward a theory of discounted repeated games with imperfect monitoring,” *Econometrica*, vol. 58, no. 5, pp. 1041–1063, 1990.
- [17] D. Fudenberg, D. K. Levine, and E. Maskin, “The folk theorem with imperfect public information,” *Econometrica*, vol. 62, no. 5, pp. 997–1039, 1994.
- [18] J. Hörner, T. Sugaya, S. Takahashi, and N. Vielle, “Recursive methods in discounted stochastic games: An algorithm for  $\delta \rightarrow 1$  and a folk theorem,” *Econometrica*, vol. 79, no. 4, pp. 1277–1318, 2011.
- [19] G. Mailath and L. Samuelson, *Repeated Games and Reputations: Long-run Relationships*. Oxford, U.K.: Oxford University Press, 2006.

## APPENDIX A

### PROOF OF PROPOSITION 1

#### A. The Claim to Prove

In order to prove Proposition 1, we quantify the performance loss of strategies restricted to  $A^{\text{as}}$ . The performance loss is determined in the following claim:

**Claim:** Starting from any initial rating profile  $\theta$ , the maximum social welfare achievable at the PAE by  $(\pi_0, \pi \cdot \mathbf{1}_N) \in \Pi(A^{\text{as}}) \times \Pi^N(A^{\text{as}})$  is at most

$$b - c - c \cdot \rho(\theta, \alpha_0^*, S_B^*) \sum_{s' \in S_B^*} q(s' | \theta, \alpha_0^*, \alpha^a \cdot \mathbf{1}_N), \quad (15)$$

where  $\alpha_0^*$ , the optimal recommended plan, and  $S_B^*$ , the optimal subset of rating distributions, are the solutions to the following optimization problem:

$$\begin{aligned} \min_{\alpha_0} \min_{S_B \subset S} & \left\{ \rho(\theta, \alpha_0, S_B) \sum_{s' \in S_B} q(s' | \theta, \alpha_0, \alpha^a \cdot \mathbf{1}_N) \right\} \\ \text{s.t.} & \sum_{s' \in S \setminus S_B} q(s' | \theta, \alpha_0, \alpha^a \cdot \mathbf{1}_N) > \sum_{s' \in S \setminus S_B} q(s' | \theta, \alpha_0, \alpha_i = \alpha^0, \alpha^a \cdot \mathbf{1}_N), \quad \forall i \in \mathcal{N}, \\ & \sum_{s' \in S \setminus S_B} q(s' | \theta, \alpha_0, \alpha^a \cdot \mathbf{1}_N) > \sum_{s' \in S \setminus S_B} q(s' | \theta, \alpha_0, \alpha_i = \alpha^1, \alpha^a \cdot \mathbf{1}_N), \quad \forall i \in \mathcal{N}, \\ & \sum_{s' \in S \setminus S_B} q(s' | \theta, \alpha_0, \alpha^a \cdot \mathbf{1}_N) > \sum_{s' \in S \setminus S_B} q(s' | \theta, \alpha_0, \alpha_i = \alpha^{01}, \alpha^a \cdot \mathbf{1}_N), \quad \forall i \in \mathcal{N}, \end{aligned} \quad (16)$$

where  $\rho(\theta, \alpha_0, S_B)$  is defined as

$$\rho(\theta, \alpha_0, S_B) \triangleq \max_{i \in \mathcal{N}} \max \left\{ \frac{\frac{s_{\theta_i} - 1}{N-1}}{\sum_{s' \in S \setminus S_B} q(s' | \theta, \alpha_0, \alpha^a \cdot \mathbf{1}_N) - q(s' | \theta, \alpha_0, \alpha_i = \alpha^0, \alpha^a \cdot \mathbf{1}_N)}, \right. \\ \left. \frac{\frac{s_{1-\theta_i}}{N-1}}{\sum_{s' \in S \setminus S_B} q(s' | \theta, \alpha_0, \alpha^a \cdot \mathbf{1}_N) - q(s' | \theta, \alpha_0, \alpha_i = \alpha^1, \alpha^a \cdot \mathbf{1}_N)}, \right. \\ \left. \frac{1}{\sum_{s' \in S \setminus S_B} q(s' | \theta, \alpha_0, \alpha^a \cdot \mathbf{1}_N) - q(s' | \theta, \alpha_0, \alpha_i = \alpha^{01}, \alpha^a \cdot \mathbf{1}_N)} \right\}, \quad (17)$$

where  $\alpha^0$  (resp.  $\alpha^1$ ) is the plan in which the user does not serve rating-0 (resp. rating-1) users, and  $\alpha^{01}$  is the plan in which the user does not serve anyone.

The above claim shows that

$$W(\varepsilon, \delta, A^{\text{as}}) \leq b - c - c \cdot \rho(\theta, \alpha_0^*, S_B^*) \sum_{s' \in S_B^*} q(s' | \theta, \alpha_0^*, \alpha^a \cdot \mathbf{1}_N)$$

for any  $\delta$ . By defining

$$\zeta(\varepsilon) \triangleq c \cdot \rho(\theta, \alpha_0^*, S_B^*) \sum_{s' \in S_B^*} q(s' | \theta, \alpha_0^*, \alpha^a \cdot \mathbf{1}_N),$$

we obtain the result in Proposition 1, namely  $\lim_{\delta \rightarrow 1} W(\varepsilon, \delta, A^{\text{as}}) \leq b - c - \zeta(\varepsilon)$ . Note that  $\zeta(\varepsilon)$  is indeed a function of the rating update error  $\varepsilon$ , because  $\varepsilon$  determines the state transition

function  $q(\mathbf{s}'|\boldsymbol{\theta}, \alpha_0^*, \alpha^a \cdot \mathbf{1}_N)$ , and thus affects  $\rho(\boldsymbol{\theta}, \alpha_0, S_B)$ . Note also that  $\zeta(\varepsilon)$  is independent of the discount factor  $\delta$ .

In the expression of  $\zeta(\varepsilon)$ ,  $\rho(\boldsymbol{\theta}, \alpha_0, S_B)$  represents the normalized benefit from deviation (normalized by  $b - c$ ). The numerator of  $\rho(\boldsymbol{\theta}, \alpha_0, S_B)$  is the probability of a player matched to the type of clients whom it deviates to not serve. The higher this probability, the larger benefit from deviation a player can get. The denominator of  $\rho(\boldsymbol{\theta}, \alpha_0, S_B)$  is the difference between the two state transition probabilities when the player does and does not deviate, respectively. When the above two transition probabilities are closer, it is less likely to detect the deviation, which results in a larger  $\rho(\boldsymbol{\theta}, \alpha_0, S_B)$ . Hence, we can expect that a larger  $\rho(\boldsymbol{\theta}, \alpha_0, S_B)$  (i.e. a larger benefit from deviation) will result in a larger performance loss, which is indeed true as will be proved later.

We can also see that  $\zeta(\varepsilon) > 0$  as long as  $\varepsilon > 0$ . The reason is as follows. Suppose that  $\varepsilon > 0$ . First, from (17), we know that  $\rho(\boldsymbol{\theta}, \alpha_0, S_B) > 0$  for any  $\boldsymbol{\theta}$ ,  $\alpha$ , and  $S_B \neq \emptyset$ . Second, we can see that  $\sum_{\mathbf{s}' \in S_B^*} q(\mathbf{s}'|\boldsymbol{\theta}, \alpha_0^*, \alpha^a \cdot \mathbf{1}_N) > 0$  as long as  $S_B^* \neq \emptyset$ . Since  $S_B^* = \emptyset$  cannot be the solution to the optimization problem (16) (because  $S_B^* = \emptyset$  violates the constraints), we know that  $\zeta(\varepsilon) > 0$ .

### B. Proof of the Claim

We prove that for any self-generating set  $(\mathcal{W}^\theta)_{\theta \in \Theta^N}$ , the maximum payoff in  $(\mathcal{W}^\theta)_{\theta \in \Theta^N}$ , namely  $\max_{\theta \in \Theta^N} \max_{v \in \mathcal{W}^\theta} \max_{i \in \mathcal{N}} v_i$ , is bounded away from the social optimum  $b - c$ , regardless of the discount factor. In this way, we can prove that any equilibrium payoff is bounded away from the social optimum. In addition, we analytically quantify the efficiency loss, which is independent of the discount factor.

Since the strategies are restricted on the subset of plans  $A^{\text{as}}$ , in each period, all the users will receive the same stage-game payoff, either  $(b - c)$  or 0, regardless of the matching rule and the rating profile. Hence, the expected discounted average payoff for each user is the same. More precisely, at any given history  $\mathbf{h}^t = (\boldsymbol{\theta}^0, \dots, \boldsymbol{\theta}^t)$ , we have

$$U_i(\boldsymbol{\theta}^t, \pi_0|\mathbf{h}^t, \pi|\mathbf{h}^t \cdot \mathbf{1}_N) = U_j(\boldsymbol{\theta}^t, \pi_0|\mathbf{h}^t, \pi|\mathbf{h}^t \cdot \mathbf{1}_N), \quad \forall i, j \in \mathcal{N}, \quad (18)$$

for any  $(\pi_0, \pi \cdot \mathbf{1}_N) \in \Pi(A^{\text{as}}) \times \Pi^N(A^{\text{as}})$ . As a result, when we restrict to the plan set  $A^{\text{as}}$ , the self-generating set  $(\mathcal{W}^\theta)_{\theta \in \Theta^N}$  satisfies for any  $\boldsymbol{\theta}$  and any  $v \in \mathcal{W}^\theta$

$$v_i = v_j, \quad \forall i, j \in \mathcal{N}. \quad (19)$$

Given any self-generating set  $(\mathcal{W}^\theta)_{\theta \in \Theta^N}$ , define the maximum payoff  $\bar{v}$  as

$$\bar{v} \triangleq \max_{\theta \in \Theta^N} \max_{v \in \mathcal{W}^\theta} \max_{i \in \mathcal{N}} v_i. \quad (20)$$

Now we derive the upper bound of  $\bar{v}$  by looking at the decomposability constraints.

To decompose the payoff profile  $\bar{v} \cdot \mathbf{1}_N$ , we must find a recommended plan  $\alpha_0 \in A^{\text{as}}$ , a plan profile  $\alpha \cdot \mathbf{1}_N$  with  $\alpha \in A^{\text{as}}$ , and a continuation payoff function  $\gamma : \Theta^N \rightarrow \cup_{\theta' \in \Theta^N} \mathcal{W}^{\theta'}$  with  $\gamma(\theta') \in \mathcal{W}^{\theta'}$ , such that for all  $i \in \mathcal{N}$  and for all  $\alpha_i \in A$ ,

$$\begin{aligned} \bar{v} &= (1 - \delta)u_i(\boldsymbol{\theta}, \alpha_0, \alpha \cdot \mathbf{1}_N) + \delta \sum_{\theta'} \gamma_i(\theta')q(\theta'|\boldsymbol{\theta}, \alpha_0, \alpha \cdot \mathbf{1}_N) \\ &\geq (1 - \delta)u_i(\boldsymbol{\theta}, \alpha_0, \alpha_i, \alpha \cdot \mathbf{1}_{N-1}) + \delta \sum_{\theta'} \gamma_i(\theta')q(\theta'|\boldsymbol{\theta}, \alpha_0, \alpha_i, \alpha \cdot \mathbf{1}_{N-1}). \end{aligned} \quad (21)$$

Note that we do *not* require the users' plan  $\alpha$  to be the same as the recommended plan  $\alpha_0$ , and that we also do *not* require the continuation payoff function  $\gamma$  to be a simple continuation payoff function.

First, the payoff profile  $\bar{v} \cdot \mathbf{1}_N$  cannot be decomposed by a recommended plan  $\alpha_0$  and the selfish plan  $\alpha^s$ . Otherwise, since  $\gamma(\theta') \in \mathcal{W}^{\theta'}$ , we have

$$\bar{v} = (1 - \delta) \cdot 0 + \delta \sum_{\theta'} \gamma_i(\theta')q(\theta'|\boldsymbol{\theta}, \alpha_0, \alpha^s \cdot \mathbf{1}_N) \leq \delta \sum_{\theta'} \bar{v}_i \cdot q(\theta'|\boldsymbol{\theta}, \alpha_0, \alpha^s \cdot \mathbf{1}_N) = \delta \cdot \bar{v} < \bar{v},$$

which is a contradiction.

Since we must use a recommended plan  $\alpha_0$  and the altruistic plan  $\alpha^a$  to decompose  $\bar{v} \cdot \mathbf{1}_N$ , we can rewrite the decomposability constraint as

$$\begin{aligned} \bar{v} &= (1 - \delta)(b - c) + \delta \sum_{\theta'} \gamma_i(\theta')q(\theta'|\boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N) \\ &\geq (1 - \delta)u_i(\boldsymbol{\theta}, \alpha_0, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1}) + \delta \sum_{\theta'} \gamma_i(\theta')q(\theta'|\boldsymbol{\theta}, \alpha_0, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1}). \end{aligned} \quad (22)$$

Since the continuation payoffs under different rating profiles  $\boldsymbol{\theta}, \boldsymbol{\theta}'$  that have the same rating distribution  $s(\boldsymbol{\theta}) = s(\boldsymbol{\theta}')$  are the same, namely  $\gamma(\boldsymbol{\theta}) = \gamma(\boldsymbol{\theta}')$ , the continuation payoff depends only on the rating distribution. For notational simplicity, with some abuse of notation, we write  $\gamma(s)$  as the continuation payoff when the rating distribution is  $s$ , write  $q(s'|\boldsymbol{\theta}, \alpha_0, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1})$  as the probability that the next state has a rating distribution  $s'$ , and write  $u_i(s, \alpha^a, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1})$  as the stage-game payoff when the next state has a rating distribution  $s$ . Then the decomposability

constraint can be rewritten as

$$\begin{aligned}\bar{v} &= (1 - \delta)(b - c) + \delta \sum_{\mathbf{s}'} \gamma_i(\mathbf{s}') q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N) \\ &\geq (1 - \delta) u_i(\mathbf{s}, \alpha_0, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1}) + \delta \sum_{\mathbf{s}'} \gamma_i(\mathbf{s}') q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1}).\end{aligned}\quad (23)$$

Now we focus on a subclass of continuation payoff functions, and derive the maximum payoff  $\bar{v}$  achievable under this subclass of continuation payoff functions. Later, we will prove that we cannot increase  $\bar{v}$  by choosing other continuation payoff functions. Specifically, we focus on a subclass of continuation payoff functions that satisfy

$$\gamma_i(\mathbf{s}) = x_A, \quad \forall i \in \mathcal{N}, \quad \forall \mathbf{s} \in S_A \subset S, \quad (24)$$

$$\gamma_i(\mathbf{s}) = x_B, \quad \forall i \in \mathcal{N}, \quad \forall \mathbf{s} \in S_B \subset S, \quad (25)$$

where  $S_A$  and  $S_B$  are subsets of the set of rating distributions  $S$  that have no intersection, namely  $S_A \cap S_B = \emptyset$ . In other words, we assign the two continuation payoff values to two subsets of rating distributions, respectively. Without loss of generality, we assume  $x_A \geq x_B$ .

Now we derive the incentive compatibility constraints. There are three plans to deviate to, the plan  $\alpha^0$  in which the user does not serve users with rating 0, the plan  $\alpha^1$  in which the user does not serve users with rating 1, and the plan  $\alpha^{01}$  in which the user does not serve anyone. The corresponding incentive compatibility constraints for a user  $i$  with rating  $\theta_i = 1$  are

$$\begin{aligned}\left[ \sum_{\mathbf{s}' \in S_A} q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N) - q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha_i = \alpha^0, \alpha^a \cdot \mathbf{1}_N) \right] (x_A - x_B) &\geq \frac{1 - \delta}{\delta} \frac{s_0}{N - 1} c, \\ \left[ \sum_{\mathbf{s}' \in S_A} q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N) - q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha_i = \alpha^1, \alpha^a \cdot \mathbf{1}_N) \right] (x_A - x_B) &\geq \frac{1 - \delta}{\delta} \frac{s_1 - 1}{N - 1} c, \\ \left[ \sum_{\mathbf{s}' \in S_A} q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N) - q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha_i = \alpha^{01}, \alpha^a \cdot \mathbf{1}_N) \right] (x_A - x_B) &\geq \frac{1 - \delta}{\delta} c.\end{aligned}\quad (26)$$

Similarly, the corresponding incentive compatibility constraints for a user  $j$  with rating  $\theta_j = 0$

are

$$\begin{aligned}
\left[ \sum_{\mathbf{s}' \in S_A} q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N) - q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha_j = \alpha^0, \alpha^a \cdot \mathbf{1}_N) \right] (x_A - x_B) &\geq \frac{1 - \delta}{\delta} \frac{s_0 - 1}{N - 1} c, \\
\left[ \sum_{\mathbf{s}' \in S_A} q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N) - q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha_j = \alpha^1, \alpha^a \cdot \mathbf{1}_N) \right] (x_A - x_B) &\geq \frac{1 - \delta}{\delta} \frac{s_1}{N - 1} c, \\
\left[ \sum_{\mathbf{s}' \in S_A} q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N) - q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha_j = \alpha^{01}, \alpha^a \cdot \mathbf{1}_N) \right] (x_A - x_B) &\geq \frac{1 - \delta}{\delta} c. \quad (27)
\end{aligned}$$

We can summarize the above incentive compatibility constraints as

$$x_A - x_B \geq \frac{1 - \delta}{\delta} c \cdot \rho(\boldsymbol{\theta}, \alpha_0, S_A), \quad (28)$$

where

$$\rho(\boldsymbol{\theta}, \alpha_0, S_B) \triangleq \max_{i \in \mathcal{N}} \max \left\{ \frac{\frac{s_{\theta_i} - 1}{N - 1}}{\sum_{\mathbf{s}' \in S \setminus S_B} q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N) - q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha_i = \alpha^0, \alpha^a \cdot \mathbf{1}_N)}, \quad (29)$$

$$\frac{\frac{s_{1 - \theta_i}}{N - 1}}{\sum_{\mathbf{s}' \in S \setminus S_B} q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N) - q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha_i = \alpha^1, \alpha^a \cdot \mathbf{1}_N)}, \quad (30)$$

$$\left. \frac{1}{\sum_{\mathbf{s}' \in S \setminus S_B} q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N) - q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha_i = \alpha^{01}, \alpha^a \cdot \mathbf{1}_N)} \right\}. \quad (31)$$

Since the maximum payoff  $\bar{v}$  satisfies

$$\bar{v} = (1 - \delta)(b - c) + \delta \left( x_A \sum_{\mathbf{s}' \in S \setminus S_B} q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N) + x_B \sum_{\mathbf{s}' \in S_B} q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N) \right), \quad (32)$$

to maximize  $\bar{v}$ , we choose  $x_B = x_A - \frac{1 - \delta}{\delta} c \cdot \rho(\boldsymbol{\theta}, \alpha_0, S_B)$ . Since  $x_A \geq \bar{v}$ , we have

$$\bar{v} = (1 - \delta)(b - c) + \delta \left( x_A - \frac{1 - \delta}{\delta} c \cdot \rho(\boldsymbol{\theta}, \alpha_0, S_B) \sum_{\mathbf{s}' \in S_B} q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N) \right) \quad (33)$$

$$\leq (1 - \delta)(b - c) + \delta \left( \bar{v} - \frac{1 - \delta}{\delta} c \cdot \rho(\boldsymbol{\theta}, \alpha_0, S_B) \sum_{\mathbf{s}' \in S_B} q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N) \right), \quad (34)$$

which leads to

$$\bar{v} \leq b - c - c \cdot \rho(\boldsymbol{\theta}, \alpha_0, S_B) \sum_{\mathbf{s}' \in S_B} q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N). \quad (35)$$

Hence, the maximum payoff  $\bar{v}$  satisfies

$$\bar{v} \leq b - c - c \cdot \min_{S_B \subset S} \left\{ \rho(\boldsymbol{\theta}, \alpha_0, S_B) \sum_{\mathbf{s}' \in S_B} q(\mathbf{s}' | \boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N) \right\}, \quad (36)$$

where  $S_B$  satisfies for all  $i \in \mathcal{N}$ ,

$$\begin{aligned} \sum_{s' \in S \setminus S_B} q(s' | \boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N) &> \sum_{s' \in S \setminus S_B} q(s' | \boldsymbol{\theta}, \alpha_0, \alpha_i = \alpha^0, \alpha^a \cdot \mathbf{1}_N), \\ \sum_{s' \in S \setminus S_B} q(s' | \boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N) &> \sum_{s' \in S \setminus S_B} q(s' | \boldsymbol{\theta}, \alpha_0, \alpha_i = \alpha^1, \alpha^a \cdot \mathbf{1}_N), \\ \sum_{s' \in S \setminus S_B} q(s' | \boldsymbol{\theta}, \alpha_0, \alpha^a \cdot \mathbf{1}_N) &> \sum_{s' \in S \setminus S_B} q(s' | \boldsymbol{\theta}, \alpha_0, \alpha_i = \alpha^{01}, \alpha^a \cdot \mathbf{1}_N). \end{aligned} \quad (37)$$

Following the same logic as in the proof of Proposition 6 in [6], we can prove that we cannot achieve a higher maximum payoff by other continuation payoff functions.

## APPENDIX B

### ANALYTICAL EXPRESSION OF $\underline{\delta}(\varepsilon, \xi)$

The lower-bound discount factor  $\underline{\delta}(\varepsilon, \xi)$  is the maximum of three critical discount factors, namely  $\underline{\delta}(\varepsilon, \xi) \triangleq \max\{\delta_1(\varepsilon, \xi), \delta_2(\varepsilon, \xi), \delta_3(\varepsilon, \xi)\}$ , where

$$\delta_1(\varepsilon, \xi) \triangleq \max_{\theta \in \{0,1\}} \frac{c}{c + (1 - 2\varepsilon)(\beta_\theta^+ - (1 - \beta_\theta^-))(\xi \frac{\kappa_2}{\kappa_1 + \kappa_2})},$$

$$\delta_2(\varepsilon, \xi) \triangleq \max_{s_1 \in \{1, \dots, N-1\}: \frac{s_1}{N-1}b + \frac{N-s_1}{N-1}c > \xi \frac{\kappa_2}{\kappa_1 + \kappa_2}} \left\{ \frac{\xi \frac{\kappa_2}{\kappa_1 + \kappa_2} - \left(\frac{s_1}{N-1}b + \frac{N-s_1}{N-1}c\right)}{\left(\xi \frac{\kappa_2}{\kappa_1 + \kappa_2}\right)(x_{s_1}^+ - x_0^+) - \left(\frac{s_1}{N-1}b + \frac{N-s_1}{N-1}c\right)} \right\},$$

and

$$\delta_3(\varepsilon, \xi) \triangleq \max_{\theta \in \{0,1\}} \frac{b - c + c \frac{x_\theta^+}{(1-2\varepsilon)[\beta_\theta^+ - (1-\beta_\theta^-)]}}{b - c + \frac{c \cdot x_\theta^+}{(1-2\varepsilon)[\beta_\theta^+ - (1-\beta_\theta^-)]} - \frac{(1+\kappa_1)(\xi \frac{\kappa_2}{\kappa_1 + \kappa_2}) - z_2}{\kappa_1} - z_3}, \quad (38)$$

where  $z_2 \triangleq -\kappa_1(b - c) + \kappa_1(1 - 1/\kappa_2)\xi + \xi \frac{\kappa_1}{\kappa_2}/(1 + \frac{\kappa_2}{\kappa_1})$ , and  $z_3 \triangleq z_2/(\kappa_1 + \kappa_2)$ . Note that  $\frac{(1+\kappa_1)(\xi \frac{\kappa_2}{\kappa_1 + \kappa_2}) - z_2}{\kappa_1} + z_3 < 0$ . We can see from the above expressions that  $\delta_1(\varepsilon, \xi) < 1$  and  $\delta_2(\varepsilon, \xi) < 1$  as long as  $\xi > 0$ . For  $\delta_3(\varepsilon, \xi)$ , simple calculations tell us that  $\xi$  appears in the denominator in the form of  $-\frac{(2\kappa_1 + \kappa_2)\kappa_2}{(\kappa_1 + \kappa_2)^2 \kappa_1} \cdot \xi$ . Since  $\kappa_1 > 0$  and  $\kappa_2 > 0$ , we know that  $-\frac{(2\kappa_1 + \kappa_2)\kappa_2}{(\kappa_1 + \kappa_2)^2 \kappa_1} < 0$ . Hence,  $\delta_3(\varepsilon, \xi)$  is increasing in  $\xi$ . As a result,  $\delta_3(\varepsilon, \xi) < 1$  as long as  $\xi$  is small enough.

Note that all the critical discount factors can be calculated analytically. Specifically,  $\delta_1(\varepsilon, \xi)$  and  $\delta_3(\varepsilon, \xi)$  are the maximum of two analytically-computed numbers, and  $\delta_2(\varepsilon, \xi)$  is the maximum of at most  $N - 1$  analytically-computed numbers.



APPENDIX C  
PROOF OF THEOREM 1

A. *Outline of the proof*

We derive the conditions under which the set  $(\mathcal{W}^\theta)_{\theta \in \Theta^N}$  is a self-generating set. Specifically, we derive the conditions under which any payoff profile  $\mathbf{v} \in \mathcal{W}^\theta$  is decomposable on  $(\mathcal{W}^{\theta'})_{\theta' \in \Theta^N}$  given  $\theta$ , for all  $\theta \in \Theta^N$ .

B. *When users have different ratings*

1) *Preliminaries:* We first focus on the states  $\theta$  with  $1 \leq s_1(\theta) \leq N - 1$ , and derive the conditions under which any payoff profile  $\mathbf{v} \in \mathcal{W}^\theta$  can be decomposed by  $(\alpha_0 = \alpha^a, \alpha^a \cdot \mathbf{1}_N)$  or  $(\alpha_0 = \alpha^f, \alpha^f \cdot \mathbf{1}_N)$ . First,  $\mathbf{v}$  could be decomposed by  $(\alpha^a, \alpha^a \cdot \mathbf{1}_N)$ , if there exists a continuation payoff function  $\gamma : \Theta^N \rightarrow \cup_{\theta' \in \Theta^N} \mathcal{W}^{\theta'}$  with  $\gamma(\theta') \in \mathcal{W}^{\theta'}$ , such that for all  $i \in \mathcal{N}$  and for all  $\alpha_i \in A$ ,

$$\begin{aligned} v_i &= (1 - \delta)u_i(\theta, \alpha^a, \alpha^a \cdot \mathbf{1}_N) + \delta \sum_{\theta'} \gamma_i(\theta')q(\theta'|\theta, \alpha^a, \alpha^a \cdot \mathbf{1}_N) \\ &\geq (1 - \delta)u_i(\theta, \alpha^a, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1}) + \delta \sum_{\theta'} \gamma_i(\theta')q(\theta'|\theta, \alpha^a, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1}). \end{aligned} \quad (39)$$

Since we focus on simple continuation payoff functions, all the users with the same future rating will have the same continuation payoff regardless of the recommended plan  $\alpha_0$ , the plan profile  $(\alpha_i, \alpha \cdot \mathbf{1}_{N-1})$ , and the future state  $\theta'$ . Hence, we write the continuation payoffs for the users with future rating 1 and 0 as  $\gamma^1$  and  $\gamma^0$ , respectively. Consequently, the above conditions on decomposability can be simplified to

$$\begin{aligned} v_i &= (1 - \delta) \cdot u_i(\theta, \alpha^a, \alpha^a \cdot \mathbf{1}_N) \\ &+ \delta \left( \gamma^1 \sum_{\theta': \theta'_i=1} q(\theta'|\theta, \alpha^a, \alpha^a \cdot \mathbf{1}_N) + \gamma^0 \sum_{\theta': \theta'_i=0} q(\theta'|\theta, \alpha^a, \alpha^a \cdot \mathbf{1}_N) \right) \\ &\geq (1 - \delta) \cdot u_i(\theta, \alpha^a, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1}) \\ &+ \delta \left( \gamma^1 \sum_{\theta': \theta'_i=1} q(\theta'|\theta, \alpha^a, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1}) + \gamma^0 \sum_{\theta': \theta'_i=0} q(\theta'|\theta, \alpha^a, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1}) \right). \end{aligned} \quad (40)$$

First, consider the case when user  $i$  has rating 1 (i.e.  $\theta_i = 1$ ). Based on (??), we can calculate the stage-game payoff as  $u_i(\theta, \alpha^a, \alpha^a \cdot \mathbf{1}_N) = b - c$ . The term  $\sum_{\theta': \theta'_i=1} q(\theta'|\theta, \alpha^a, \alpha^a \cdot \mathbf{1}_N)$  is the

probability that user  $i$  has rating 1 in the next state. Since user  $i$ 's rating update is independent of the other users' rating update, we can calculate this probability as

$$\sum_{\theta':\theta'_i=1} q(\theta'|\theta, \alpha^a, \alpha^a \cdot \mathbf{1}_N) = [(1-\varepsilon)\beta_1^+ + \varepsilon(1-\beta_1^-)] \sum_{m \in M:\theta_{m(i)}=1} \mu(m) \quad (41)$$

$$+ [(1-\varepsilon)\beta_1^+ + \varepsilon(1-\beta_1^-)] \sum_{m \in M:\theta_{m(i)}=0} \mu(m) \quad (42)$$

$$= (1-\varepsilon)\beta_1^+ + \varepsilon(1-\beta_1^-) = x_1^+. \quad (43)$$

Similarly, we can calculate  $\sum_{\theta':\theta'_i=0} q(\theta'|\theta, \alpha^a, \alpha^a \cdot \mathbf{1}_N)$ , the probability that user  $i$  has rating 0 in the next state, as

$$\sum_{\theta':\theta'_i=0} q(\theta'|\theta, \alpha^a, \alpha^a \cdot \mathbf{1}_N) = [(1-\varepsilon)(1-\beta_1^+) + \varepsilon\beta_1^-] \sum_{m \in M:\theta_{m(i)}=1} \mu(m) \quad (44)$$

$$+ [(1-\varepsilon)(1-\beta_1^+) + \varepsilon\beta_1^-] \sum_{m \in M:\theta_{m(i)}=0} \mu(m) \quad (45)$$

$$= (1-\varepsilon)(1-\beta_1^+) + \varepsilon\beta_1^- = 1 - x_1^+. \quad (46)$$

Now we discuss what happens if user  $i$  deviates. Since the recommended plan  $\alpha^a$  is to exert high effort for all the users, user  $i$  can deviate to the other three plans, namely “exert high effort for rating-1 users only”, “exert high effort for rating-0 users only”, “exert low effort for all the users”. We can calculate the corresponding stage-game payoff and state transition probabilities under each deviation.

- “exert high effort for rating-1 users only” ( $\alpha_i(1, \theta_i) = 1, \alpha_i(0, \theta_i) = 0$ ):

$$u_i(\theta, \alpha^a, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1}) = b - c \cdot \sum_{m \in M:\theta_{m(i)}=1} \mu(m) = b - c \cdot \frac{s_1(\theta) - 1}{N-1} \quad (47)$$

$$\begin{aligned} & \sum_{\theta':\theta'_i=1} q(\theta'|\theta, \alpha^a, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1}) \quad (48) \\ &= [(1-\varepsilon)\beta_1^+ + \varepsilon(1-\beta_1^-)] \sum_{m \in M:\theta_{m(i)}=1} \mu(m) + [(1-\varepsilon)(1-\beta_1^-) + \varepsilon\beta_1^+] \sum_{m \in M:\theta_{m(i)}=0} \mu(m) \\ &= [(1-\varepsilon)\beta_1^+ + \varepsilon(1-\beta_1^-)] \frac{s_1(\theta) - 1}{N-1} + [(1-\varepsilon)(1-\beta_1^-) + \varepsilon\beta_1^+] \frac{s_0(\theta)}{N-1}. \end{aligned}$$

$$\begin{aligned}
& \sum_{\boldsymbol{\theta}': \theta'_i=0} q(\boldsymbol{\theta}'|\boldsymbol{\theta}, \alpha^a, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1}) \tag{49} \\
&= [(1-\varepsilon)(1-\beta_1^+) + \varepsilon\beta_1^-] \sum_{m \in M: \theta_{m(i)}=1} \mu(m) + [(1-\varepsilon)\beta_1^- + \varepsilon(1-\beta_1^+)] \sum_{m \in M: \theta_{m(i)}=0} \mu(m) \\
&= [(1-\varepsilon)(1-\beta_1^+) + \varepsilon\beta_1^-] \frac{s_1(\boldsymbol{\theta}) - 1}{N-1} + [(1-\varepsilon)\beta_1^- + \varepsilon(1-\beta_1^+)] \frac{s_0(\boldsymbol{\theta})}{N-1}.
\end{aligned}$$

- “exert high effort for rating-0 users only” ( $\alpha_i(1, \theta_i) = 0, \alpha_i(0, \theta_i) = 1$ ):

$$u_i(\boldsymbol{\theta}, \alpha^a, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1}) = b - c \cdot \sum_{m \in M: \theta_{m(i)}=0} \mu(m) = b - c \cdot \frac{s_0(\boldsymbol{\theta})}{N-1} \tag{50}$$

$$\begin{aligned}
& \sum_{\boldsymbol{\theta}': \theta'_i=1} q(\boldsymbol{\theta}'|\boldsymbol{\theta}, \alpha^a, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1}) \tag{51} \\
&= [(1-\varepsilon)(1-\beta_1^-) + \varepsilon\beta_1^+] \sum_{m \in M: \theta_{m(i)}=1} \mu(m) + [(1-\varepsilon)\beta_1^+ + \varepsilon(1-\beta_1^-)] \sum_{m \in M: \theta_{m(i)}=0} \mu(m) \\
&= [(1-\varepsilon)(1-\beta_1^-) + \varepsilon\beta_1^+] \frac{s_1(\boldsymbol{\theta}) - 1}{N-1} + [(1-\varepsilon)\beta_1^+ + \varepsilon(1-\beta_1^-)] \frac{s_0(\boldsymbol{\theta})}{N-1}.
\end{aligned}$$

$$\begin{aligned}
& \sum_{\boldsymbol{\theta}': \theta'_i=0} q(\boldsymbol{\theta}'|\boldsymbol{\theta}, \alpha^a, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1}) \tag{52} \\
&= [(1-\varepsilon)\beta_1^- + \varepsilon(1-\beta_1^+)] \sum_{m \in M: \theta_{m(i)}=1} \mu(m) + [(1-\varepsilon)(1-\beta_1^+) + \varepsilon\beta_1^-] \sum_{m \in M: \theta_{m(i)}=0} \mu(m) \\
&= [(1-\varepsilon)\beta_1^- + \varepsilon(1-\beta_1^+)] \frac{s_1(\boldsymbol{\theta}) - 1}{N-1} + [(1-\varepsilon)(1-\beta_1^+) + \varepsilon\beta_1^-] \frac{s_0(\boldsymbol{\theta})}{N-1}.
\end{aligned}$$

- “exert low effort for all the users” ( $\alpha_i(1, \theta_i) = 0, \alpha_i(0, \theta_i) = 0$ ):

$$u_i(\boldsymbol{\theta}, \alpha^a, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1}) = b \tag{53}$$

$$\begin{aligned}
& \sum_{\boldsymbol{\theta}': \theta'_i=1} q(\boldsymbol{\theta}'|\boldsymbol{\theta}, \alpha^a, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1}) \tag{54} \\
&= [(1-\varepsilon)(1-\beta_1^-) + \varepsilon\beta_1^+] \sum_{m \in M: \theta_{m(i)}=1} \mu(m) + [(1-\varepsilon)(1-\beta_1^-) + \varepsilon\beta_1^+] \sum_{m \in M: \theta_{m(i)}=0} \mu(m) \\
&= (1-\varepsilon)(1-\beta_1^-) + \varepsilon\beta_1^+.
\end{aligned}$$

$$\begin{aligned}
& \sum_{\boldsymbol{\theta}': \theta'_i=0} q(\boldsymbol{\theta}' | \boldsymbol{\theta}, \alpha^a, \alpha_i, \alpha^a \cdot \mathbf{1}_{N-1}) \\
&= [(1-\varepsilon)\beta_1^- + \varepsilon(1-\beta_1^+)] \sum_{m \in M: \theta_{m(i)}=1} \mu(m) + [(1-\varepsilon)\beta_1^- + \varepsilon(1-\beta_1^+)] \sum_{m \in M: \theta_{m(i)}=0} \mu(m) \\
&= (1-\varepsilon)\beta_1^- + \varepsilon(1-\beta_1^+).
\end{aligned} \tag{55}$$

Plugging the above expressions into (40), we can simplify the incentive compatibility constraints (i.e. the inequality constraints) to

$$(1-2\varepsilon) [\beta_1^+ - (1-\beta_1^-)] (\gamma^1 - \gamma^0) \geq \frac{1-\delta}{\delta} \cdot c, \tag{56}$$

under all three deviating plans.

Hence, if user  $i$  has rating 1, the decomposability constraints (40) reduces to

$$v^1 = (1-\delta) \cdot (b-c) + \delta \cdot [x_1^+ \gamma^1 + (1-x_1^+) \gamma^0], \tag{57}$$

where  $v^1$  is the payoff of the users with rating 1, and

$$(1-2\varepsilon) [\beta_1^+ - (1-\beta_1^-)] (\gamma^1 - \gamma^0) \geq \frac{1-\delta}{\delta} \cdot c. \tag{58}$$

Similarly, if user  $i$  has rating 0, we can reduce the decomposability constraints (40) to

$$v^0 = (1-\delta) \cdot (b-c) + \delta \cdot [x_0^+ \gamma^1 + (1-x_0^+) \gamma^0], \tag{59}$$

and

$$(1-2\varepsilon) [\beta_0^+ - (1-\beta_0^-)] (\gamma^1 - \gamma^0) \geq \frac{1-\delta}{\delta} \cdot c. \tag{60}$$

For the above incentive compatibility constraints (the above two inequalities) to hold, we need to have  $\beta_1^+ - (1-\beta_1^-) > 0$  and  $\beta_0^+ - (1-\beta_0^-) > 0$ , which are part of Condition 1 and Condition 2. Now we will derive the rest of the sufficient conditions in Theorem 1.

The above two equalities determine the continuation payoff  $\gamma^1$  and  $\gamma^0$  as below

$$\begin{cases} \gamma^1 = \frac{1}{\delta} \cdot \frac{(1-x_0^+)v^1 - (1-x_1^+)v^0}{x_1^+ - x_0^+} - \frac{1-\delta}{\delta} \cdot (b-c) \\ \gamma^0 = \frac{1}{\delta} \cdot \frac{x_1^+v^0 - x_0^+v^1}{x_1^+ - x_0^+} - \frac{1-\delta}{\delta} \cdot (b-c) \end{cases}. \tag{61}$$

Now we consider the decomposability constraints if we want to decompose a payoff profile  $\mathbf{v} \in \mathcal{W}^\theta$  using the fair plan  $\alpha^f$ . Since we focus on decomposition by simple continuation payoff

functions, we write the decomposition constraints as

$$\begin{aligned}
v_i &= (1 - \delta) \cdot u_i(\boldsymbol{\theta}, \alpha^f, \alpha^f \cdot \mathbf{1}_N) \\
&+ \delta \left( \gamma^1 \sum_{\boldsymbol{\theta}': \theta'_i=1} q(\boldsymbol{\theta}' | \boldsymbol{\theta}, \alpha^f, \alpha^f \cdot \mathbf{1}_N) + \gamma^0 \sum_{\boldsymbol{\theta}': \theta'_i=0} q(\boldsymbol{\theta}' | \boldsymbol{\theta}, \alpha^f, \alpha^f \cdot \mathbf{1}_N) \right) \\
&\geq (1 - \delta) \cdot u_i(\boldsymbol{\theta}, \alpha^f, \alpha_i, \alpha^f \cdot \mathbf{1}_{N-1}) \\
&+ \delta \left( \gamma^1 \sum_{\boldsymbol{\theta}': \theta'_i=1} q(\boldsymbol{\theta}' | \boldsymbol{\theta}, \alpha^f, \alpha_i, \alpha^f \cdot \mathbf{1}_{N-1}) + \gamma^0 \sum_{\boldsymbol{\theta}': \theta'_i=0} q(\boldsymbol{\theta}' | \boldsymbol{\theta}, \alpha^f, \alpha_i, \alpha^f \cdot \mathbf{1}_{N-1}) \right).
\end{aligned} \tag{62}$$

Due to space limitation, we omit the details and directly give the simplification of the above decomposability constraints as follows. First, the incentive compatibility constraints (i.e. the inequality constraints) are simplified to

$$(1 - 2\varepsilon) [\beta_1^+ - (1 - \beta_1^-)] (\gamma^1 - \gamma^0) \geq \frac{1 - \delta}{\delta} \cdot c, \tag{63}$$

and

$$(1 - 2\varepsilon) [\beta_0^+ - (1 - \beta_0^-)] (\gamma^1 - \gamma^0) \geq \frac{1 - \delta}{\delta} \cdot c, \tag{64}$$

under all three deviating plans. Note that the above incentive compatibility constraints are the same as the ones when we want to decompose the payoffs using the altruistic plan  $\alpha^a$ .

Then, the equality constraints in (62) can be simplified as follows. For the users with rating 1, we have

$$v^1 = (1 - \delta) \cdot \left( b - \frac{s_1(\boldsymbol{\theta}) - 1}{N - 1} c \right) + \delta \cdot \left[ x_{s_1(\boldsymbol{\theta})}^+ \cdot \gamma^1 + (1 - x_{s_1(\boldsymbol{\theta})}^+) \cdot \gamma^0 \right], \tag{65}$$

where

$$x_{s_1(\boldsymbol{\theta})} \triangleq \left[ (1 - \varepsilon) \frac{s_1(\boldsymbol{\theta}) - 1}{N - 1} + \frac{s_0(\boldsymbol{\theta})}{N - 1} \right] \beta_1^+ + \left( \varepsilon \frac{s_1(\boldsymbol{\theta}) - 1}{N - 1} \right) (1 - \beta_1^-). \tag{66}$$

For the users with rating 0, we have

$$v^0 = (1 - \delta) \cdot \left( \frac{s_0(\boldsymbol{\theta}) - 1}{N - 1} b - c \right) + \delta \cdot \left[ x_0^+ \gamma^1 + (1 - x_0^+) \gamma^0 \right]. \tag{67}$$

The above two equalities determine the continuation payoff  $\gamma^1$  and  $\gamma^0$  as below

$$\begin{cases} \gamma^1 = \frac{1}{\delta} \cdot \frac{(1 - x_0^+) v^1 - (1 - x_{s_1(\boldsymbol{\theta})}^+) v^0}{x_{s_1(\boldsymbol{\theta})}^+ - x_0^+} - \frac{1 - \delta}{\delta} \cdot \frac{\left( b - \frac{s_1(\boldsymbol{\theta}) - 1}{N - 1} c \right) (1 - x_0^+) - \left( \frac{s_0(\boldsymbol{\theta}) - 1}{N - 1} b - c \right) (1 - x_{s_1(\boldsymbol{\theta})}^+)}{x_{s_1(\boldsymbol{\theta})}^+ - x_0^+} \\ \gamma^0 = \frac{1}{\delta} \cdot \frac{x_{s_1(\boldsymbol{\theta})}^+ v^0 - x_0^+ v^1}{x_{s_1(\boldsymbol{\theta})}^+ - x_0^+} - \frac{1 - \delta}{\delta} \cdot \frac{\left( b - \frac{s_1(\boldsymbol{\theta}) - 1}{N - 1} c \right) x_0^+ - \left( \frac{s_0(\boldsymbol{\theta}) - 1}{N - 1} b - c \right) x_{s_1(\boldsymbol{\theta})}^+}{x_{s_1(\boldsymbol{\theta})}^+ - x_0^+} \end{cases}. \tag{68}$$

2) *Sufficient conditions:* Now we derive the sufficient conditions under which any payoff profile  $\mathbf{v} \in \mathcal{W}^\theta$  can be decomposed by  $(\alpha_0 = \alpha^a, \alpha^a \cdot \mathbf{1}_N)$  or  $(\alpha_0 = \alpha^f, \alpha^f \cdot \mathbf{1}_N)$ . Specifically, we will derive the conditions such that for any payoff profile  $\mathbf{v} \in \mathcal{W}^\theta$ , at least one of the two decomposability constraints (40) and (62) is satisfied. From the preliminaries, we know that the incentive compatibility constraints in (40) and (62) can be simplified into the same constraints:

$$(1 - 2\varepsilon) [\beta_1^+ - (1 - \beta_1^-)] (\gamma^1 - \gamma^0) \geq \frac{1 - \delta}{\delta} \cdot c, \quad (69)$$

and

$$(1 - 2\varepsilon) [\beta_0^+ - (1 - \beta_0^-)] (\gamma^1 - \gamma^0) \geq \frac{1 - \delta}{\delta} \cdot c. \quad (70)$$

The above constraints impose the constraint on the discount factor, namely

$$\delta \geq \max_{\theta \in \Theta} \frac{c}{c + (1 - 2\varepsilon) [\beta_\theta^+ - (1 - \beta_\theta^-)] (\gamma^1 - \gamma^0)}. \quad (71)$$

Since  $\gamma_1$  and  $\gamma_0$  should satisfy  $\gamma^1 - \gamma^0 \geq \epsilon_0 - \epsilon_1$ , the above constraints can be rewritten as

$$\delta \geq \max_{\theta \in \Theta} \frac{c}{c + (1 - 2\varepsilon) [\beta_\theta^+ - (1 - \beta_\theta^-)] (\epsilon_0 - \epsilon_1)}, \quad (72)$$

where is part of Condition 3 in Theorem 1.

In addition, the continuation payoffs  $\gamma_1$  and  $\gamma_0$  should satisfy the constraints of the self-generating set, namely

$$\gamma^1 - \gamma^0 \geq \epsilon_0 - \epsilon_1, \quad (73)$$

$$\gamma^1 + \frac{c}{(N-1)b} \cdot \gamma^0 \leq z_2 \triangleq (1 + \frac{c}{(N-1)b})(b-c) - \frac{c}{(N-1)b} \epsilon_0 - \epsilon_1, \quad (74)$$

$$\gamma^1 - \frac{b}{\frac{N-2}{N-1}b-c} \cdot \gamma^0 \leq z_3 \triangleq -\frac{\frac{b}{\frac{N-2}{N-1}b-c} - 1}{1 + \frac{c}{(N-1)b}} \cdot z_2. \quad (75)$$

We can plug the expressions of the continuation payoffs  $\gamma_1$  and  $\gamma_0$  in (61) and (68) into the above constraints. Specifically, if a payoff profile  $\mathbf{v}$  is decomposed by the altruistic plan, the following constraints should be satisfied for the continuation payoff profile to be in the self-generating set: (for notational simplicity, we define  $\kappa_1 \triangleq \frac{b}{\frac{N-2}{N-1}b-c} - 1$  and  $\kappa_2 \triangleq 1 + \frac{c}{(N-1)b}$ )

$$\frac{1}{\delta} \cdot \frac{v^1 - v^0}{x_1^+ - x_0^+} \geq \epsilon_0 - \epsilon_1, \quad (\alpha^a-1)$$

$$\frac{1}{\delta} \cdot \left\{ \frac{(1 - \kappa_2 x_0^+)v^1 - (1 - \kappa_2 x_1^+)v^0}{x_1^+ - x_0^+} - \kappa_2 \cdot (b-c) \right\} \leq z_2 - \kappa_2 \cdot (b-c), \quad (\alpha^a-2)$$

$$\frac{1}{\delta} \cdot \left\{ \frac{(1 + \kappa_1 x_0^+)v^1 - (1 + \kappa_1 x_1^+)v^0}{x_1^+ - x_0^+} + \kappa_1 \cdot (b-c) \right\} \leq z_3 + \kappa_1 \cdot (b-c). \quad (\alpha^a-3)$$

The constraint  $(\alpha^a-1)$  is satisfied for all  $v^1$  and  $v^0$  as long as  $x_1^+ > x_0^+$ , because  $v^1 - v^0 > \epsilon_0 - \epsilon_1$ ,  $|x_1^+ > x_0^+| < 1$ , and  $\delta < 1$ .

Since both the left-hand side (LHS) and the right-hand side (RHS) of  $(\alpha^a-2)$  are smaller than 0, we have

$$(\alpha^a-2) \Leftrightarrow \delta \leq \frac{\frac{(1-\kappa_2 x_0^+)v^1 - (1-\kappa_2 x_1^+)v^0}{x_1^+ - x_0^+} - \kappa_2 \cdot (b - c)}{z_2 - \kappa_2 \cdot (b - c)} \quad (76)$$

The RHS of  $(\alpha^a-3)$  is larger than 0. Hence, we have

$$(\alpha^a-3) \Leftrightarrow \delta \geq \frac{\frac{(1+\kappa_1 x_0^+)v^1 - (1+\kappa_1 x_1^+)v^0}{x_1^+ - x_0^+} + \kappa_1 \cdot (b - c)}{z_3 + \kappa_1 \cdot (b - c)}. \quad (77)$$

If a payoff profile  $v$  is decomposed by the fair plan, the following constraints should be satisfied for the continuation payoff profile to be in the self-generating set:

$$\frac{1}{\delta} \cdot \left\{ \frac{v^1 - v^0}{x_{s_1(\theta)}^+ - x_0^+} - \frac{\frac{s_1(\theta)}{N-1}b + \frac{s_0(\theta)}{N-1}c}{x_{s_1(\theta)}^+ - x_0^+} \right\} \geq \epsilon_0 - \epsilon_1 - \frac{\frac{s_1(\theta)}{N-1}b + \frac{s_0(\theta)}{N-1}c}{x_{s_1(\theta)}^+ - x_0^+}, \quad (\alpha^f-1)$$

$$\begin{aligned} & \frac{1}{\delta} \cdot \left\{ \frac{(1 - \kappa_2 x_0^+)v^1 - (1 - \kappa_2 x_{s_1(\theta)}^+)v^0}{x_{s_1(\theta)}^+ - x_0^+} - \frac{(1 - \kappa_2 x_0^+) \left( b - \frac{s_1(\theta)-1}{N-1}c \right) - (1 - \kappa_2 x_{s_1(\theta)}^+) \left( \frac{s_0(\theta)-1}{N-1}b - c \right)}{x_{s_1(\theta)}^+ - x_0^+} \right\} \\ & \leq z_2 - \frac{(1 - \kappa_2 x_0^+) \left( b - \frac{s_1(\theta)-1}{N-1}c \right) - (1 - \kappa_2 x_{s_1(\theta)}^+) \left( \frac{s_0(\theta)-1}{N-1}b - c \right)}{x_{s_1(\theta)}^+ - x_0^+}, \quad (\alpha^f-2) \end{aligned}$$

$$\begin{aligned} & \frac{1}{\delta} \cdot \left\{ \frac{(1 + \kappa_1 x_0^+)v^1 - (1 + \kappa_1 x_{s_1(\theta)}^+)v^0}{x_{s_1(\theta)}^+ - x_0^+} - \frac{(1 + \kappa_1 x_0^+) \left( b - \frac{s_1(\theta)-1}{N-1}c \right) - (1 + \kappa_1 x_{s_1(\theta)}^+) \left( \frac{s_0(\theta)-1}{N-1}b - c \right)}{x_{s_1(\theta)}^+ - x_0^+} \right\} \\ & \leq z_3 - \frac{(1 + \kappa_1 x_0^+) \left( b - \frac{s_1(\theta)-1}{N-1}c \right) - (1 + \kappa_1 x_{s_1(\theta)}^+) \left( \frac{s_0(\theta)-1}{N-1}b - c \right)}{x_{s_1(\theta)}^+ - x_0^+}. \quad (\alpha^f-3) \end{aligned}$$

Since  $\frac{v^1 - v^0}{x_{s_1(\theta)}^+ - x_0^+} > \epsilon_0 - \epsilon_1$ , the constraint  $(\alpha^f-1)$  is satisfied for all  $v^1$  and  $v^0$  if  $v^1 - v^0 \geq \frac{s_1(\theta)}{N-1}b + \frac{s_0(\theta)}{N-1}c$ . Hence, the constraint  $(\alpha^f-1)$  is equivalent to

$$\delta \geq \frac{v^1 - v^0 - \left( \frac{s_1(\theta)}{N-1}b + \frac{s_0(\theta)}{N-1}c \right)}{(\epsilon^0 - \epsilon^1)(x_{s_1(\theta)}^+ - x_0^+) - \left( \frac{s_1(\theta)}{N-1}b + \frac{s_0(\theta)}{N-1}c \right)}, \text{ for } \theta \text{ s.t. } \frac{s_1(\theta)}{N-1}b + \frac{s_0(\theta)}{N-1}c \geq v^1 - v^0. \quad (78)$$

For  $(\alpha^f-2)$ , we want to make the RHS have the same (minus) sign under any state  $\theta$ , which is true if

$$1 - \kappa_2 x_0^+ > 0, \quad 1 - \kappa_2 x_{s_1(\theta)}^+ < 0, \quad \frac{1 - \kappa_2 x_{s_1(\theta)}^+}{1 - \kappa_2 x_0^+} \geq -(\kappa_2 - 1), \quad s_1(\theta) = 1, \dots, N-1, \quad (79)$$

which leads to

$$x_{s_1(\theta)}^+ > \frac{1}{\kappa_2}, \quad x_0^+ < \frac{1}{\kappa_2}, \quad x_0^+ < \frac{1 - x_{s_1(\theta)}^+}{1 - \kappa_2}, \quad s_1(\theta) = 1, \dots, N-1, \quad (80)$$

$$\Leftrightarrow \frac{N-2}{N-1}x_1^+ + \frac{1}{N-1}\beta_1^+ > \frac{1}{\kappa_2}, \quad x_0^+ < \min \left\{ \frac{1}{\kappa_2}, \frac{1 - \beta_1^+}{1 - \kappa_2} \right\}. \quad (81)$$

Since the RHS of  $(\alpha^f-2)$  is smaller than 0, we have

$$(\alpha^f-2) \Leftrightarrow \delta \leq \frac{\frac{(1-\kappa_2x_0^+)v^1 - (1-\kappa_2x_{s_1(\theta)}^+)v^0}{x_{s_1(\theta)}^+ - x_0^+} - \frac{(1-\kappa_2x_0^+)\left(b - \frac{s_1(\theta)-1}{N-1}c\right) - (1-\kappa_2x_{s_1(\theta)}^+)\left(\frac{s_0(\theta)-1}{N-1}b - c\right)}{x_{s_1(\theta)}^+ - x_0^+}}{z_2 - \frac{(1-\kappa_2x_0^+)\left(b - \frac{s_1(\theta)-1}{N-1}c\right) - (1-\kappa_2x_{s_1(\theta)}^+)\left(\frac{s_0(\theta)-1}{N-1}b - c\right)}{x_{s_1(\theta)}^+ - x_0^+}}. \quad (82)$$

For  $(\alpha^f-3)$ , since  $\frac{1+\kappa_1x_{s_1(\theta)}^+}{1+\kappa_1x_0^+} < 1 + \kappa_1$ , the RHS is always smaller than 0. Hence, we have

$$(\alpha^f-3) \Leftrightarrow \delta \leq \frac{\frac{(1+\kappa_1x_0^+)v^1 - (1+\kappa_1x_{s_1(\theta)}^+)v^0}{x_{s_1(\theta)}^+ - x_0^+} - \frac{(1+\kappa_1x_0^+)\left(b - \frac{s_1(\theta)-1}{N-1}c\right) - (1+\kappa_1x_{s_1(\theta)}^+)\left(\frac{s_0(\theta)-1}{N-1}b - c\right)}{x_{s_1(\theta)}^+ - x_0^+}}{z_3 - \frac{(1+\kappa_1x_0^+)\left(b - \frac{s_1(\theta)-1}{N-1}c\right) - (1+\kappa_1x_{s_1(\theta)}^+)\left(\frac{s_0(\theta)-1}{N-1}b - c\right)}{x_{s_1(\theta)}^+ - x_0^+}}. \quad (83)$$

We briefly summarize what requirements on  $\delta$  we have obtained now. To make the continuation payoff profile in the self-generating under the decomposition of  $\alpha^a$ , we have one upper bound on  $\delta$  resulting from  $(\alpha^a-2)$  and one lower bound on  $\delta$  resulting from  $(\alpha^a-3)$ . To make the continuation payoff profile in the self-generating under the decomposition of  $\alpha^f$ , we have two upper bounds on  $\delta$  resulting from  $(\alpha^f-2)$  and  $(\alpha^f-3)$ , and one lower bound on  $\delta$  resulting from  $(\alpha^f-1)$ . First, we want to eliminate the upper bounds, namely make the upper bounds larger than 1, such that  $\delta$  can be arbitrarily close to 1.

To eliminate the following upper bound resulting from  $(\alpha^a-2)$

$$\delta \leq \frac{\frac{(1-\kappa_2x_0^+)v^1 - (1-\kappa_2x_1^+)v^0}{x_1^+ - x_0^+} - \kappa_2 \cdot (b - c)}{z_2 - \kappa_2 \cdot (b - c)}, \quad (84)$$

we need to have (since  $z_2 - \kappa_2 \cdot (b - c) < 0$ )

$$\frac{(1 - \kappa_2x_0^+)v^1 - (1 - \kappa_2x_1^+)v^0}{x_1^+ - x_0^+} \leq z_2, \quad \forall v^1, v^0. \quad (85)$$

The LHS of the above inequality is maximized when  $v^0 = \frac{z_2 - z_3}{\kappa_1 + \kappa_2}$  and  $v^1 = v^0 + \frac{\kappa_1 z_2 + \kappa_2 z_3}{\kappa_1 + \kappa_2}$ . Hence, the above inequality is satisfied if

$$\frac{(1 - \kappa_2x_0^+) \left( \frac{z_2 - z_3}{\kappa_1 + \kappa_2} + \frac{\kappa_1 z_2 + \kappa_2 z_3}{\kappa_1 + \kappa_2} \right) - (1 - \kappa_2x_1^+) \frac{z_2 - z_3}{\kappa_1 + \kappa_2}}{x_1^+ - x_0^+} \leq z_2 \quad (86)$$

$$\Leftrightarrow \left( \frac{1 - x_1^+ - x_0^+(\kappa_2 - 1)}{x_1^+ - x_0^+} \frac{\kappa_1}{\kappa_1 + \kappa_2} \right) z_2 \leq - \frac{1 - x_1^+ - x_0^+(\kappa_2 - 1)}{x_1^+ - x_0^+} \frac{\kappa_2}{\kappa_1 + \kappa_2} z_3. \quad (87)$$



Since  $x_0^+ < \frac{1-\beta_1^+}{1-\kappa_2} < \frac{1-x_1^+}{1-\kappa_2}$ , we have

$$z_2 \leq -\frac{\kappa_2}{\kappa_1} z_3. \quad (88)$$

To eliminate the following upper bound resulting from ( $\alpha^f$ -2)

$$\delta \leq \frac{\frac{(1-\kappa_2 x_0^+)v^1 - (1-\kappa_2 x_{s_1(\theta)}^+)v^0}{x_{s_1(\theta)}^+ - x_0^+} - \frac{(1-\kappa_2 x_0^+)\left(b - \frac{s_1(\theta)-1}{N-1}c\right) - (1-\kappa_2 x_{s_1(\theta)}^+)\left(\frac{s_0(\theta)-1}{N-1}b - c\right)}{x_{s_1(\theta)}^+ - x_0^+}}{z_2 - \frac{(1-\kappa_2 x_0^+)\left(b - \frac{s_1(\theta)-1}{N-1}c\right) - (1-\kappa_2 x_{s_1(\theta)}^+)\left(\frac{s_0(\theta)-1}{N-1}b - c\right)}{x_{s_1(\theta)}^+ - x_0^+}}, \quad (89)$$

we need to have (since  $z_2 - \frac{(1-\kappa_2 x_0^+)\left(b - \frac{s_1(\theta)-1}{N-1}c\right) - (1-\kappa_2 x_{s_1(\theta)}^+)\left(\frac{s_0(\theta)-1}{N-1}b - c\right)}{x_{s_1(\theta)}^+ - x_0^+} < 0$ )

$$\frac{(1 - \kappa_2 x_0^+)v^1 - (1 - \kappa_2 x_{s_1(\theta)}^+)v^0}{x_{s_1(\theta)}^+ - x_0^+} \leq z_2, \quad \forall v^1, v^0. \quad (90)$$

Similarly, the LHS of the above inequality is maximized when  $v^0 = \frac{z_2 - z_3}{\kappa_1 + \kappa_2}$  and  $v^1 = v^0 + \frac{\kappa_1 z_2 + \kappa_2 z_3}{\kappa_1 + \kappa_2}$ . Hence, the above inequality is satisfied if

$$\frac{(1 - \kappa_2 x_0^+) \left( \frac{z_2 - z_3}{\kappa_1 + \kappa_2} + \frac{\kappa_1 z_2 + \kappa_2 z_3}{\kappa_1 + \kappa_2} \right) - (1 - \kappa_2 x_{s_1(\theta)}^+) \frac{z_2 - z_3}{\kappa_1 + \kappa_2}}{x_{s_1(\theta)}^+ - x_0^+} \leq z_2 \quad (91)$$

$$\Leftrightarrow \left( \frac{1 - x_{s_1(\theta)}^+ - x_0^+(\kappa_2 - 1)}{x_{s_1(\theta)}^+ - x_0^+} \frac{\kappa_1}{\kappa_1 + \kappa_2} \right) z_2 \leq -\frac{1 - x_{s_1(\theta)}^+ - x_0^+(\kappa_2 - 1)}{x_{s_1(\theta)}^+ - x_0^+} \frac{\kappa_2}{\kappa_1 + \kappa_2} z_3. \quad (92)$$

Since  $x_0^+ < \frac{1-\beta_1^+}{1-\kappa_2} < \frac{1-x_{s_1(\theta)}^+}{1-\kappa_2}$ , we have

$$z_2 \leq -\frac{\kappa_2}{\kappa_1} z_3. \quad (93)$$

To eliminate the following upper bound resulting from ( $\alpha^f$ -3)

$$\delta \leq \frac{\frac{(1+\kappa_1 x_0^+)v^1 - (1+\kappa_1 x_{s_1(\theta)}^+)v^0}{x_{s_1(\theta)}^+ - x_0^+} - \frac{(1+\kappa_1 x_0^+)\left(b - \frac{s_1(\theta)-1}{N-1}c\right) - (1+\kappa_1 x_{s_1(\theta)}^+)\left(\frac{s_0(\theta)-1}{N-1}b - c\right)}{x_{s_1(\theta)}^+ - x_0^+}}{z_3 - \frac{(1+\kappa_1 x_0^+)\left(b - \frac{s_1(\theta)-1}{N-1}c\right) - (1+\kappa_1 x_{s_1(\theta)}^+)\left(\frac{s_0(\theta)-1}{N-1}b - c\right)}{x_{s_1(\theta)}^+ - x_0^+}}, \quad (94)$$

we need to have (since  $z_3 - \frac{(1+\kappa_1 x_0^+)\left(b - \frac{s_1(\theta)-1}{N-1}c\right) - (1+\kappa_1 x_{s_1(\theta)}^+)\left(\frac{s_0(\theta)-1}{N-1}b - c\right)}{x_{s_1(\theta)}^+ - x_0^+} < 0$ )

$$\frac{(1 + \kappa_1 x_0^+)v^1 - (1 + \kappa_1 x_{s_1(\theta)}^+)v^0}{x_{s_1(\theta)}^+ - x_0^+} \leq z_3, \quad \forall v^1, v^0. \quad (95)$$

Again, the LHS of the above inequality is maximized when  $v^0 = \frac{z_2 - z_3}{\kappa_1 + \kappa_2}$  and  $v^1 = v^0 + \frac{\kappa_1 z_2 + \kappa_2 z_3}{\kappa_1 + \kappa_2}$ .

Hence, the above inequality is satisfied if

$$\frac{(1 + \kappa_1 x_0^+) \left( \frac{z_2 - z_3}{\kappa_1 + \kappa_2} + \frac{\kappa_1 z_2 + \kappa_2 z_3}{\kappa_1 + \kappa_2} \right) - (1 + \kappa_1 x_{s_1(\theta)}^+) \frac{z_2 - z_3}{\kappa_1 + \kappa_2}}{x_{s_1(\theta)}^+ - x_0^+} \leq z_3 \quad (96)$$

$$\Leftrightarrow \left( \frac{1 - x_{s_1(\theta)}^+ + x_0^+(\kappa_1 + 1)}{x_{s_1(\theta)}^+ - x_0^+} \frac{\kappa_1}{\kappa_1 + \kappa_2} \right) z_2 \leq - \frac{1 - x_{s_1(\theta)}^+ + x_0^+(\kappa_1 + 1)}{x_{s_1(\theta)}^+ - x_0^+} \frac{\kappa_2}{\kappa_1 + \kappa_2} z_3. \quad (97)$$

Since  $1 - x_{s_1(\theta)}^+ + x_0^+(\kappa_1 + 1) > 0$ , we have

$$z_2 \leq - \frac{\kappa_2}{\kappa_1} z_3. \quad (98)$$

In summary, to eliminate the upper bounds on  $\delta$ , we only need to have  $z_2 \leq - \frac{\kappa_2}{\kappa_1} z_3$ , which is satisfied since we define  $z_3 \triangleq - \frac{\kappa_1}{\kappa_2} z_2$ .

Now we derive the analytical lower bound on  $\delta$  based on the lower bounds resulting from  $(\alpha^a-3)$  and  $(\alpha^f-1)$ :

$$(\alpha^a-3) \Leftrightarrow \delta \geq \frac{\frac{(1 + \kappa_1 x_0^+) v^1 - (1 + \kappa_1 x_1^+) v^0}{x_1^+ - x_0^+} + \kappa_1 \cdot (b - c)}{z_3 + \kappa_1 \cdot (b - c)}, \quad (99)$$

and

$$\delta \geq \frac{v^1 - v^0 - \left( \frac{s_1(\theta)}{N-1} b + \frac{s_0(\theta)}{N-1} c \right)}{(\epsilon^0 - \epsilon^1)(x_{s_1(\theta)}^+ - x_0^+) - \left( \frac{s_1(\theta)}{N-1} b + \frac{s_0(\theta)}{N-1} c \right)}, \text{ for } \theta \text{ s.t. } \frac{s_1(\theta)}{N-1} b + \frac{s_0(\theta)}{N-1} c \geq v^1 - v^0. \quad (100)$$

We define an intermediate lower bound based on the latter inequality along with the inequalities resulting from the incentive compatibility constraints:

$$\underline{\delta}' = \max \left\{ \max_{s_1 \in \{1, \dots, N-1\}: \frac{s_1-1}{N-1} b + \frac{N-s_1}{N-1} c > \epsilon_0 - \epsilon_1} \frac{\epsilon_0 - \epsilon_1 - \left( \frac{s_1}{N-1} b + \frac{N-s_1}{N-1} c \right)}{(\epsilon_0 - \epsilon_1) \left( \frac{N-s_1}{N-1} \beta_1^+ + \frac{s_1-1}{N-1} x_1^+ \right) - \left( \frac{s_1}{N-1} b + \frac{N-s_1}{N-1} c \right)}, \max_{\theta \in \{0,1\}} \frac{c}{c + (1 - 2\varepsilon)(\beta_\theta^+ - (1 - \beta_\theta^-))(\epsilon_0 - \epsilon_1)} \right\} \quad (101)$$

Then the lower bound can be written as  $\underline{\delta} = \max \{ \underline{\delta}', \underline{\delta}'' \}$ , where  $\underline{\delta}''$  is the lower bound that we will derive for the case when the users have the same rating. If the payoffs  $v^1$  and  $v^0$  satisfy the constraint resulting from  $(\alpha^a-3)$ , namely satisfy

$$\frac{(1 + \kappa_1 x_0^+) v^1 - (1 + \kappa_1 x_1^+) v^0}{x_1^+ - x_0^+} \leq \underline{\delta} z_3 - (1 - \underline{\delta}) \kappa_1 \cdot (b - c), \quad (102)$$

then we use  $\alpha^a$  to decompose  $v^1$  and  $v^0$ . Otherwise, we use  $\alpha^f$  to decompose  $v^1$  and  $v^0$

### C. When the users have the same rating

Now we derive the conditions under which any payoff profile in  $\mathcal{W}^{1_N}$  and  $\mathcal{W}^{0_N}$  can be decomposed.

If all the users have rating 1, namely  $\boldsymbol{\theta} = \mathbf{1}_N$ , to decompose  $\mathbf{v} \in \mathcal{W}^{1_N}$ , we need to find a recommended plan  $\alpha_0$  and a simple continuation payoff function  $\gamma$  such that for all  $i \in \mathcal{N}$  and for all  $\alpha_i \in A$ ,

$$\begin{aligned} v_i &= (1 - \delta)u_i(\boldsymbol{\theta}, \alpha_0, \alpha_0 \cdot \mathbf{1}_N) + \delta \sum_{\boldsymbol{\theta}'} \gamma_i(\boldsymbol{\theta}')q(\boldsymbol{\theta}'|\boldsymbol{\theta}, \alpha_0, \alpha_0 \cdot \mathbf{1}_N) \\ &\geq (1 - \delta)u_i(\boldsymbol{\theta}, \alpha_0, \alpha_i, \alpha_0 \cdot \mathbf{1}_{N-1}) + \delta \sum_{\boldsymbol{\theta}'} \gamma_i(\boldsymbol{\theta}')q(\boldsymbol{\theta}'|\boldsymbol{\theta}, \alpha_0, \alpha_i, \alpha_0 \cdot \mathbf{1}_{N-1}). \end{aligned} \quad (103)$$

When all the users have the same rating, the altruistic plan  $\alpha^a$  is equivalent to the fair plan  $\alpha^f$ . Hence, we use the altruistic plan and the selfish plan to decompose the payoff profiles.

If we use the altruistic plan  $\alpha^a$  to decompose a payoff profile  $\mathbf{v}$ , we have

$$v^1 = (1 - \delta)(b - c) + \delta(x_1^+ \gamma^1 + (1 - x_1^+) \gamma^0), \quad (104)$$

and the incentive compatibility constraint

$$(1 - 2\varepsilon) [\beta_1^+ - (1 - \beta_1^-)] (\gamma^1 - \gamma^0) \geq \frac{1 - \delta}{\delta} c. \quad (105)$$

Setting  $\gamma^1 = \gamma^0 + \frac{1 - \delta}{\delta} \frac{c}{(1 - 2\varepsilon) [\beta_1^+ - (1 - \beta_1^-)]}$  and noticing that  $\gamma^0 \in \left[ \frac{(1 + \kappa_1)(\epsilon_0 - \epsilon_1) - z_3}{\kappa_1}, \frac{\kappa_1 z_2 + (\kappa_2 - 1)z_3}{\kappa_1 + \kappa_2} \right]$ , we get an lower bound on  $v^1$  that can be decomposed by  $\alpha^a$

$$v^1 = (1 - \delta)(b - c) + \delta \left( \gamma^0 + x_1^+ \frac{1 - \delta}{\delta} \frac{c}{(1 - 2\varepsilon) [\beta_1^+ - (1 - \beta_1^-)]} \right) \quad (106)$$

$$\geq (1 - \delta) \left( b - c + c \frac{x_1^+}{(1 - 2\varepsilon) [\beta_1^+ - (1 - \beta_1^-)]} \right) + \delta \frac{(1 + \kappa_1)(\epsilon_0 - \epsilon_1) - z_3}{\kappa_1} \quad (107)$$

If we use the selfish plan  $\alpha^s$  to decompose a payoff profile  $\mathbf{v}$ , we have

$$v^1 = \delta(x_1^+ \gamma^1 + (1 - x_1^+) \gamma^0). \quad (108)$$

Since the selfish plan is NE of the stage game, the incentive compatibility constraint is satisfied as long as we set  $\gamma^1 = \gamma^0$ . Hence, we have  $v^1 = \delta\gamma^0$ . Again, noticing that  $\gamma^0 \in \left[ \frac{(1 + \kappa_1)(\epsilon_0 - \epsilon_1) - z_3}{\kappa_1}, \frac{\kappa_1 z_2 + (\kappa_2 - 1)z_3}{\kappa_1 + \kappa_2} \right]$ , we get an upper bound on  $v^1$  that can be decomposed by  $\alpha^s$

$$v^1 = \delta\gamma^0 \leq \delta \frac{\kappa_1 z_2 + (\kappa_2 - 1)z_3}{\kappa_1 + \kappa_2}. \quad (109)$$

In order to decompose any payoff profile  $\mathbf{v} \in \mathcal{W}^{1N}$ , the lower bound on  $v^1$  that can be decomposed by  $\alpha^a$  must be smaller than the upper bound on  $v^1$  that can be decomposed by  $\alpha^s$ , which leads to

$$(1 - \delta) \left( b - c + c \frac{x_1^+}{(1-2\varepsilon)[\beta_1^+ - (1-\beta_1^-)]} \right) + \delta \frac{(1+\kappa_1)(\varepsilon_0 - \varepsilon_1) - z_3}{\kappa_1} \leq \delta \frac{\kappa_1 z_2 + (\kappa_2 - 1)z_3}{\kappa_1 + \kappa_2}$$

$$\Rightarrow \delta \geq \frac{b - c + c \frac{x_1^+}{(1-2\varepsilon)[\beta_1^+ - (1-\beta_1^-)]}}{b - c + c \frac{x_1^+}{(1-2\varepsilon)[\beta_1^+ - (1-\beta_1^-)]} + \frac{\kappa_1 z_2 + (\kappa_2 - 1)z_3}{\kappa_1 + \kappa_2} - \frac{(1+\kappa_1)(\varepsilon_0 - \varepsilon_1) - z_3}{\kappa_1}}. \quad (110)$$

Finally, following the same procedure, we derive the lower bound on  $\delta$  when all the users have rating 0, namely  $\boldsymbol{\theta} = \mathbf{0}_N$ . Similarly, in this case, the altruistic plan  $\alpha^a$  is equivalent to the fair plan  $\alpha^f$ . Hence, we use the altruistic plan and the selfish plan to decompose the payoff profiles.

If we use the altruistic plan  $\alpha^a$  to decompose a payoff profile  $\mathbf{v}$ , we have

$$v^0 = (1 - \delta)(b - c) + \delta(x_0^+ \gamma^1 + (1 - x_0^+) \gamma^0), \quad (111)$$

and the incentive compatibility constraint

$$(1 - 2\varepsilon) [\beta_0^+ - (1 - \beta_0^-)] (\gamma^1 - \gamma^0) \geq \frac{1 - \delta}{\delta} c. \quad (112)$$

If we use the selfish plan  $\alpha^s$  to decompose a payoff profile  $\mathbf{v}$ , we have

$$v^1 = \delta(x_0^+ \gamma^1 + (1 - x_0^+) \gamma^0). \quad (113)$$

Note that when  $\boldsymbol{\theta} = \mathbf{0}_N$ , if we substitute  $\beta_0^+$ ,  $\beta_0^-$ ,  $x_0^-$  with  $\beta_1^+$ ,  $\beta_1^-$ ,  $x_1^-$ , respectively, the decomposability constraints become the same as those when  $\boldsymbol{\theta} = \mathbf{1}_N$ . Hence, we derive a similar lower bound on  $\delta$

$$\delta \geq \frac{b - c + c \frac{x_0^+}{(1-2\varepsilon)[\beta_0^+ - (1-\beta_0^-)]}}{b - c + c \frac{x_0^+}{(1-2\varepsilon)[\beta_0^+ - (1-\beta_0^-)]} + \frac{\kappa_1 z_2 + (\kappa_2 - 1)z_3}{\kappa_1 + \kappa_2} - \frac{(1+\kappa_1)(\varepsilon_0 - \varepsilon_1) - z_3}{\kappa_1}}. \quad (114)$$

Finally, we can obtain the lower bound on  $\delta$  when the users have the same rating as

$$\underline{\delta}'' = \max_{\theta \in \{0,1\}} \frac{b - c + c \frac{x_\theta^+}{(1-2\varepsilon)[\beta_\theta^+ - (1-\beta_\theta^-)]}}{b - c + c \frac{x_\theta^+}{(1-2\varepsilon)[\beta_\theta^+ - (1-\beta_\theta^-)]} + \frac{\kappa_1 z_2 + (\kappa_2 - 1)z_3}{\kappa_1 + \kappa_2} - \frac{(1+\kappa_1)(\varepsilon_0 - \varepsilon_1) - z_3}{\kappa_1}}. \quad (115)$$

Together with the lower bound  $\underline{\delta}'$  derived for the case when the users have different ratings, we can get the lower bound  $\underline{\delta}$  specified in Condition 3 of Theorem 1.

APPENDIX D  
COMPLETE DESCRIPTION OF THE ALGORITHM

TABLE VII. THE ALGORITHM OF CONSTRUCTING THE EQUILIBRIUM STRATEGY BY THE RATING MECHANISM.

<b>Require:</b> $b, c, \varepsilon, \xi; \tau(\varepsilon), \delta \geq \underline{\delta}(\varepsilon, \xi); \theta^0$	<i>(inputs to the algorithm)</i>
<b>Initialization:</b> $t = 0, \epsilon_0 = \xi, \epsilon_1 = \epsilon_0 / (1 + \frac{\kappa_2}{\kappa_1}), v^\theta = b - c - \epsilon_\theta, \theta = \theta^0.$	<i>(set the target payoffs)</i>
<b>repeat</b>	
<b>if</b> $s_1(\theta) = 0$ <b>then</b>	
<b>if</b> $v^0 \geq (1 - \delta) \left[ b - c + \frac{(1-\varepsilon)\beta_0^+ + \varepsilon(1-\beta_0^-)}{(1-2\varepsilon)(\beta_0^+ - (1-\beta_0^-))} c \right] + \delta \frac{\epsilon_0 - \epsilon_1 - z_3}{\kappa_1}$ <b>then</b>	
$\alpha_0^t = \alpha^a$	<i>(determine the recommended plan)</i>
$v^0 \leftarrow \frac{v^0}{\delta} - \frac{1-\delta}{\delta} \left[ b - c + \frac{(1-\varepsilon)\beta_0^+ + \varepsilon(1-\beta_0^-)}{(1-2\varepsilon)(\beta_0^+ - (1-\beta_0^-))} c \right], v^1 \leftarrow v^0 + \frac{1-\delta}{\delta} \left[ \frac{1}{(1-2\varepsilon)(\beta_0^+ - (1-\beta_0^-))} c \right]$	<i>(update the continuation payoff)</i>
<b>else</b>	
$\alpha_0^t = \alpha^s$	<i>(determine the recommended plan)</i>
$v^0 \leftarrow \frac{v^0}{\delta}, v^1 \leftarrow v^0$	<i>(update the continuation payoff)</i>
<b>end</b>	
<b>elseif</b> $s_1(\theta) = N$ <b>then</b>	
<b>if</b> $v^1 \geq (1 - \delta) \left[ b - c + \frac{(1-\varepsilon)\beta_1^+ + \varepsilon(1-\beta_1^-)}{(1-2\varepsilon)(\beta_1^+ - (1-\beta_1^-))} c \right] + \delta \frac{\epsilon_0 - \epsilon_1 - z_3}{\kappa_1}$ <b>then</b>	
$\alpha_0^t = \alpha^a$	<i>(determine the recommended plan)</i>
$v^1 \leftarrow \frac{v^1}{\delta} - \frac{1-\delta}{\delta} \left[ b - c + \frac{(1-\varepsilon)\beta_1^+ + \varepsilon(1-\beta_1^-)}{(1-2\varepsilon)(\beta_1^+ - (1-\beta_1^-))} c \right], v^0 \leftarrow v^1 - \frac{1-\delta}{\delta} \left[ \frac{1}{(1-2\varepsilon)(\beta_1^+ - (1-\beta_1^-))} c \right]$	<i>(update the continuation payoff)</i>
<b>else</b>	
$\alpha_0^t = \alpha^s$	<i>(determine the recommended plan)</i>
$v^1 \leftarrow \frac{v^1}{\delta}, v^0 \leftarrow v^1$	<i>(update the continuation payoff)</i>
<b>end</b>	
<b>else</b>	
<b>if</b> $\frac{1+\kappa_1 x_0^+}{x_1^+ - x_0^+} v^1 - \frac{1+\kappa_1 x_1^+}{x_1^+ - x_0^+} v^0 \leq \delta z_3 - (1 - \delta)\kappa_1(b - c)$ <b>then</b>	
$\alpha_0^t = \alpha^a$	<i>(determine the recommended plan)</i>
$v^{1'} \leftarrow \frac{1}{\delta} \frac{(1-x_0^+)v^1 - (1-x_1^+)v^0}{x_1^+ - x_0^+} - \frac{1-\delta}{\delta} (b - c), v^{0'} \leftarrow \frac{1}{\delta} \frac{x_1^+ v^0 - x_0^+ v^1}{x_1^+ - x_0^+} - \frac{1-\delta}{\delta} (b - c)$	<i>(update the continuation payoff)</i>
$v^1 \leftarrow v^{1'}, v^0 \leftarrow v^{0'}$	
<b>else</b>	
$\alpha_0^t = \alpha^f$	<i>(determine the recommended plan)</i>
$v^{1'} \leftarrow \frac{1}{\delta} \frac{(1-x_0^+)v^1 - (1-x_{s_1(\theta)}^+)v^0}{x_{s_1(\theta)}^+ - x_0^+} - \frac{1-\delta}{\delta} \frac{(b - \frac{s_1(\theta)-1}{N-1}c)(1-x_0^+) - (\frac{s_0(\theta)-1}{N-1}b - c)(1-x_{s_1(\theta)}^+)}{x_{s_1(\theta)}^+ - x_0^+}$	<i>(update the continuation payoff)</i>
$v^{1'} \leftarrow \frac{1}{\delta} \frac{x_0^+ v^1 - x_{s_1(\theta)}^+ v^0}{x_0^+ - x_{s_1(\theta)}^+} - \frac{1-\delta}{\delta} \frac{(b - \frac{s_1(\theta)-1}{N-1}c)x_0^+ - (\frac{s_0(\theta)-1}{N-1}b - c)x_{s_1(\theta)}^+}{x_0^+ - x_{s_1(\theta)}^+}$	
$v^1 \leftarrow v^{1'}, v^0 \leftarrow v^{0'}$	
<b>end</b>	
<b>end</b>	
$t \leftarrow t + 1$ , determine the rating profile $\theta^t$ , set $\theta \leftarrow \theta^t$	
<b>until</b> $\emptyset$	