Skill Management in Large-scale Service Marketplaces

Gad Allon
The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania, gadallon@wharton.upenn.edu

Achal Bassamboo
Kellogg School of Management, Northwestern University, Evanston, Illinois, a-bassamboo@kellogg.northwestern.edu

Eren B. Çil
Lundquist College of Business, University of Oregon, Eugene, Oregon, erencil@uoregon.edu

Large-scale, web-based service marketplaces have recently emerged as a new resource for customers who need quick resolutions for their short-term problems. Due to the temporary nature of the relations between customers and service providers (agents) in these marketplaces, customers may not have an opportunity to assess the ability of an agent before their service completion. On the other hand, the moderating firm has a more sustained relationship with agents, and thus it can provide customers with more information about the abilities of agents through skill screening mechanisms. In this paper, we consider a marketplace where the moderating firm can run two skills tests on agents to assess if their skills are above certain thresholds. Our main objective is to evaluate the effectiveness of skill screening as a revenue maximization tool. We, specifically, analyze how much benefit the firm obtains after each additional skill test. We find that skill screening leads to negligible revenue improvements in marketplaces where agent skills are highly compatible and the average service times are similar for all customers. As the compatibility of agent skills weakens or the customers start to vary in their processing time needs, we show that the firm starts to experience sizable improvements in revenue from skill screening. Apparently, the firm can reap the most of these substantial benefits when it runs only one test. For instance, in marketplaces where agents posses uncorrelated skills, the second skill test only brings an additional 2% improvement in revenue. Accounting for possible skill screening costs, we then show the optimality of offering only one test when the compatibility between agent skills is sufficiently low. The results of this paper also have important implications in terms of the right level of intervention in the marketplaces we study.

Key words: Service marketplaces; skill management; flexible resources; price competition; non-cooperative game theory.

History: Received: September 2015; accepted: June 2017 by Michael Pinedo after three revisions.

1. Introduction

Large-scale, web-based service marketplaces have recently emerged as a new resource for customers who need quick resolutions for their temporary problems. In these marketplaces, many small service providers (agents) compete among themselves to help customers with diverse needs. Typically, an independent firm, which we shall refer to as the *moderating firm*, establishes the infrastructure for the interaction between customers and agents in these marketplaces. In particular, the moderating firm provides the customers and the agents with the information required to make their decisions. A notable example among many existing online marketplaces is upwork.com (formerly odesk.com). The web-site hosts around 9,000,000 programmers competing to provide software solutions.

Considering large-scale natures of online marketplaces, it is not surprising to see that the ability of agents to serve customers with a particular need varies significantly. Naturally, customers prefer to be serviced by a more skilled agent because a more capable agent is likely to generate more value for customers. Unfortunately, customers may not have an opportunity to assess the ability of an agent before their service completion because most of the relations between customers and providers are temporary in these marketplaces. On the other hand, the moderating firm has a more sustained relationship with agents, and thus it can obtain more information about their abilities. Particularly, the firm can constitute a *skill screening mechanism*. In general, these mechanisms take the form of skill tests and/or certification programs that are run by moderating firms. For instance, upwork.com offers various exams to test the ability of the *candidate* providers. In fact, being successful in some of these exams is the first requirement for providers to be eligible to serve customers in the marketplace. The makers of upwork.com (or any other moderating firm) freely decide on how comprehensive the exams are. The more comprehensive the exams become, the more value customers expect from their service. If necessary, upwork.com can use these exams to disqualify some of the agents, and thus control the portfolio of different agent types (e.g., flexible, dedicated) and the service capacity in the marketplace. We use the term *skill-mix structure* to denote the portfolio of different agent types. Most of the online service marketplaces, including upwork.com, receives 10% of the revenue obtained by the agents at service completion. Therefore, it is in the best interest of the moderating firm to intervene in the marketplace by using its skill tests in order to make sure that the “right” prices and customer demand emerge in the marketplace.

Motivated by these online service marketplaces, we consider a marketplace with two groups of customers, each of which has different needs. In real life, customers may seek help on web programming, graphic design, translation, proofreading, video production and editing, etc. in online service marketplaces. We use the term *class* to identify the group of customers with the same needs. On the supply side, we assume there is a finite but large number of agents who are homogeneous in their service capacity and heterogeneous in the value that their services generate for each customer class. Specifically, the service of an agent generates a random value (with a known distribution) for each customer class. We refer to these two random values as *agent skills*. Customers cannot observe agent skills, but the moderating firm can run two skills tests on agents to assess whether their skills are above certain thresholds or not. The firm determines a passing level in each skill test and allows agents to serve customers only if their skills are above the passing levels. We use a game theoretical framework to study the interaction between the customers and the agents. Namely, each agent announces a price for his service, and customers request service from agents based on the expected agent skills, the prices, and the anticipated waiting time. The objective of the firm is to find the passing levels that maximize its revenue, which is a predetermined share of the total revenue generated in the marketplace.

As we mention above, the moderating firm can use its screening mechanism as a tool to influence the revenue generated in the marketplace. In this paper, we aim at evaluating the effectiveness of skill screening as a revenue maximization tool. To this end, we analyze how much benefit the firm obtains after each additional skill test. We start with a benchmark case where the firm does not offer any skill tests. Then, we study the firm's problem (i) when the firm uses only one exam, and (ii) when both exams are offered. Due to test preparations and executions, moderating firms may incur a cost to carry out the skill tests. Taking the possibility of costly skill screening into consideration, we compare the revenues in these three cases to find the optimal number of skill tests. One may view the number of tests offered by the firm as a measure for how strictly the firm regulates the marketplace. Thus, our analysis also provides insights for the right level of intervention in the marketplaces we study.

In analyzing the model described above, we observe that the optimization problem of the firm becomes analytically intractable when the skills of an agent follow a general joint probability distribution. Thus, we obtain the firm's optimal decisions by considering a family of skill distributions with a shape parameter which controls the correlation between agent skills. Since the online marketplaces we review usually house service providers with compatible skills, especially in programming, we focus on positively correlated skills in this paper.

Our model of agent skills enables us to gain a better understanding of the relationship between skill correlation and how the firm utilizes skill screening. We, specifically, study how the revenue benefits from skill screening depend on the correlation between the agent skills. We find that skill screening leads to minimal revenue improvements in marketplaces where agent skills are highly correlated and the average processing times are similar for all customers. If the moderating firm is concerned about the cost of screening, this result suggests that the firm is better off not offering any skill tests when the skill correlation is high. As the compatibility of agent skills weakens, we show that the firm starts to experience substantial revenue benefits from skill screening. Particularly, we prove that the firm's benefit from skill screening can be as much as 25% if agent skills exhibit negligible correlation and customers require the same service time on average. We also show that the skill screening becomes more effective as the different classes of customers start to vary in terms of their processing time needs. Apparently, the firm can reap the most of these substantial benefits when it runs only one exam. For instance, in marketplaces with almost independent skills, the second skill test can only bring an additional 2% revenue improvement. As a matter of fact, the firm does not gain any benefits from the second exam in markets where the customer demand is above a critical level. Accounting for possible testing costs, we then show the optimality of offering only one test when skill correlation is sufficiently low.

The results of this paper also have important implications in terms of the moderating firm's involvements in the marketplace. Our findings suggest that the firm does not need to regulate the marketplace via skill screening when agents are endowed with highly compatible skills. When intervention is needed, we establish that it is sufficient to run only one of the exams as an intervention tool when considering costly skill screening. The contribution of this paper is also in introducing a family of joint skill distributions that captures the correlation between skills ranging from perfect and positive correlation to no correlation. Our methodology can be easily extended to study the service environments with negatively correlated skills.

2. Literature Review

Our paper lies in the intersection of various streams of research. The first line of works related to our paper studies customer behavior in service systems. Service systems with customers who seek to maximize their utilities have attracted the attention of researchers for many years. The analysis of such systems dates back to Naor's seminal work (See Naor (1969)), which analyzes customer behavior in a single-server queueing system. More recently, Cachon and Harker (2002) and Allon and Federgruen (2007) study the competition between multiple firms offering substitute but differentiated services by modeling the customer behavior implicitly via an exogenously given demand function. An alternative approach is followed in Chen and Wan (2003), where authors examine the customers' choice problem explicitly by embedding it into the firms' pricing problem.

Our paper is also related to the research focusing on the economic trade-offs between investing on flexible resources, which provide the ability to satisfy a wide variety of customer needs, and dedicated resources responding to only a specific demand type. This line of literature studies a two-stage decision problem with recourse, which is also known as the Newsvendor Network problem, and dates back to Fine and Freund (1990). Fine and Freund (1990) considers a firm that invests in a portfolio of multiple dedicated resources and one flexible resource in the first stage where the market demand for its products is uncertain. After making the capacity investments, the demand uncertainty is resolved, and the firm makes the production decisions to maximize its profit. Fine and Freund (1990) argues that the flexible resource is not preferred when demand distributions are perfectly and positively correlated. Gupta et al. (1992) studies a similar model where the firm initially has some existing capacity and presents results parallel to Fine and Freund (1990). Contrary to the examples provided in these two papers, Callen and Sarath (1995) and Van Mieghem (1998) show that it can be optimal for a firm to invest in a flexible resource even if demand distributions are perfectly and positively correlated. Recent papers extend the model in Fine and Freund (1990) by studying the optimal pricing decision of a monopolist (See Chod and Rudi (2005) and Bish and Wang (2004)), competition between two firms (See Goyal and Netessine (2007)), and more detailed configurations of flexibility (See Bassamboo et al. (2010)). In all of these papers, the firm chooses its price and allocates its flexible capacity in order to maximize its profit. However, in the service marketplaces we consider, the (moderating) firm does not have direct control over the pricing and the service decisions of the service providers.

The pricing and the capacity planning problem of the service systems can easily become analytically intractable when trying to study more complex models, such as a multi-server queueing systems. Recognizing this difficulty, Halfin and Whitt (1981) proposes a framework to obtain robust and accurate approximations to analyze multi-server queues. This framework has been applied by many researchers to study the pricing and service design problem of a monopoly in more realistic and detailed settings. Armony and Maglaras (2004) and Maglaras and Zeevi (2003) are notable examples of papers using the asymptotic analysis to tackle complexity of these problems. Furthermore, Zeltyn and Mandelbaum (2005) extends the asymptotic analysis of markovian queueing system by considering customer abandonments.

The idea of using approximation methods can also be applied to characterize the equilibrium behavior of the firms in a competitive environment. In this paper, we study the game between customers and service providers by constructing an approximation of the original model. Our approximation is based on the fluid analysis framework introduced in Whitt (2006). To our knowledge, Allon and Gurvich (2010) and Chen et al. (2008) are the first papers studying competition among complex service systems via asymptotic analysis. There are two main differences between these two papers and our work. First, both of them study a service environment with a fixed number of decision makers (firms), while the number of decision makers in our marketplace (agents) grows unboundedly. Second, they only consider a competitive environment where the firms behave individually. In contrast, we study a marketplace where the agents have a limited level of collaboration. Another recent paper that studies the equilibrium characterization of a competitive marketplace is

Allon et al. (2012). It studies different involvements of the moderating firm in a service marketplace supposing a fixed skill-mix structure. Specifically, the moderating firm can introduce operational tools which provide an efficient match between customers and providers. Moreover, the firms can provide strategic tools which allow communication and collaboration among the agents. Allon et al. (2012) concludes that the moderating firm should compliment its operational tools by creating communication opportunities among providers. The major difference between our paper andAllon et al. (2012) is that we explore the effects of different skill-mix structures on the moderating firm's revenue, supposing the firm offers both efficient matching between customers and providers and communication among providers.

The research on marketplaces may also be viewed as related to the literature on labor markets that studies the wage dynamics (See Burdett and Mortensen (1998), Manning (2004), and Michaelides (2010)). In this paper, our focus is on a market for temporary help, which means that the engagement between customers and service providers ends upon the service completion. This stands in contrast to the labor economics literature in which the engagement is assumed to be permanent. Furthermore, the entities governing the labor markets can use intervention tools directly influencing the wage dynamic, such as minimum wage. Unlike the literature in labor markets, the moderating firms we consider have minimal direct power to influence the prices emerging as the equilibrium outcomes. Our paper also differs from the literature on market microstructure. This body of work studies market makers who can set prices and hold inventories of assets in order to stabilize markets (See Garman (1976), Amihud and Mendelson (1980), Ho and Stoll (1983), and a comprehensive survey by Biais et al. (2005)). However, the moderating firms considered in our paper have no direct price-setting power and cannot respond to customers' service requests.

3. Model Basics

Consider a service marketplace where agents and customers make their decisions in order to maximize their individual utilities. There are two groups of customers, each of which seek a different and unique service skill. We use the term "*class*" to identify the group of customers with the same service needs. We refer to one customer class as class A and the other one as class B. The service requests from class $i \in \{A, B\}$ customers follows a Poisson process with rate Λ_i. This forms the "potential demand" for the marketplace. Each customer is a risk neutral individual and thus decides whether to join the marketplace or not according to her expected utility. Customers make their decisions after observing the decisions of the agents. Customers who join the marketplace form the "effective demand" for the marketplace. If a customer decides not to join the system, she requests the service from an outside option which generates a utility of $\underline{u}$. If a class $i \in \{A, B\}$ customer joins the system, she incurs a waiting cost of c_i per unit time until her service commences. We assume that service time required to satisfy the requests of a class $i \in \{A, B\}$ customer is exponentially distributed with rate τ_i. When the service of a class customer is successfully completed, she pays the price of the service, earns a reward which depends on the skills of the agent serving her. The expected utility of a customer is based on the reward, the price, and the anticipated waiting time.

The above summarizes the demand arriving to the marketplace. Next, we discuss the capacity provision in the marketplace. There are k *candidate* agents endowed with different processing skills. Particularly, the value that an agent's service generates for a class $i \in \{A, B\}$ customer is S_i. S_A and S_B are random variables with a joint probability density function $f_{A,B}(\cdot, \cdot)$ on the support $[0, \bar{R}_A] \times [0, \bar{R}_B]$. We refer to (S_A, S_B) as the *agent skills* and $f_{A,B}(\cdot, \cdot)$ as the *skill distribution.* The skills are not observable but the moderating firm can verify whether agent skills are above a threshold through a skill screening process. As long as an agent is eligible to serve customers, he makes a pricing decision for his service. Furthermore, agents, who are qualified to serve both classes, choose how to distribute their service capacity among the customer classes. Each agent makes these decisions independently in order to maximize his expected revenue. The expected revenue of an agent depends on the price he charges and his demand volume. We normalized the operating cost of the agents to zero for notational convenience. We also suppose that agents are identical in terms of any characteristics other than the processing skills, e.g., work experience, customer satisfaction, etc., in order to evaluate the importance of skill test in a more transparent manner. It is also worth noting that in this paper, we focus our attention to the jobs that require flexible agents to commit to a service decision such as allocation of their service capacity among different customer classes. It is possible that such a service commitment may not be necessary for all jobs. The jobs that do not require service commitments can be studied in a model where flexible agents only set their prices and compete for both customers classes simultaneously. In this more complex model, we can show the existence of equilibria that are revenue equivalent to the equilibrium outcomes our paper predicts. A major analytical challenge in a model without service commitments would be to show the uniqueness of the equilibria. However, even if the lack of service commitment led to additional equilibria, our qualitative insights about the effectiveness of skill screening would continue to hold as long as the most profitable equilibrium emerges as the market outcome. Moderating firms may sustain the most profitable equilibrium as the market outcome by influencing their agents' decisions through educational/training tools like the Knowledge Center at upwork.com.

We refer to the ratio $\Lambda_i/(\tau_i k)$ as the demand-supply ratio of class $i \in \{A, B\}$ and denote it by $\rho_i > 0$. The demand-supply ratio is a first order measure for the mismatch between aggregate demand and the total processing capacity for each class. In this section, we describe only the basics of the model. Next section discusses more details about the marketplace model we study. We also present a table in Appendix A that provides the description of the frequently used notation.

4. The Roles of the Moderating Firm

The essential role of the moderating firm in a large-scale marketplace is to construct the infrastructure for the interaction between players. This is crucial because all players have to be equipped with the necessary information, such as prices to make their decisions, and individual players cannot gather this information on their own. There are also other ways for moderating firms to be involved in a marketplace. For instance, moderating firms can provide mechanisms which improve the operational performance of the whole system by efficiently matching customers and agents. They may also complement their operational tools with strategic tools which enable communication among agents. Furthermore, because agents' skills are not observable to the customers, moderating firms may provide customers with further information about the candidate agents by screening agents' abilities. In this section, our goal is to build a model where we capture these different roles of the moderating firms. To this end, we first introduce a screening mechanism which consists of skill tests determining whether a candidate agent is eligible to serve customers or not. Next, we provide a detailed description of the interaction between customers and agents in a marketplace when operational inefficiencies are minimized and agents are allowed to communicate. Using this model, we will study the moderating firm's skill and capacity management problem with the objective of revenue maximization in Section 6. Note that the firm's only source of revenue is its predetermined share of the total revenue generated in the marketplace. Therefore, the moderating firm uses its screening mechanism to maximize the total revenue in the system.

4.1. Setting up the Skill-Mix

As we mentioned in the Introduction, moderating firms can obtain more information about the abilities of candidate agents through a screening process. We model this by assuming that the moderating firm can run two skill tests on each candidate agent, say Exam A and Exam B, in order to screen his abilities. In particular, in Exam $i \in \{A, B\}$, the firm picks a threshold level ω_i (a measure for comprehensiveness) and tests whether the value that an agent's service generates for class i is above ω_i. We refer to these thresholds as the *passing levels*. As an example, at upwork.com, the passing level for most tests is 2.5 out of 5.

The firm publicly announces the results of the tests, and a candidate agent will be *eligible* to serve customers if he passes at least one exam. An agent who passes only Exam $i \in \{A, B\}$, will be eligible to serve only class i, and thus his only decision will be to set a price for his service. We refer to these types of agents as *dedicated* agents and denote the fraction of dedicated agents for class $i \in \{A, B\}$ by α_i Candidate agents who pass both exams will be eligible to serve both classes of customers. We refer to these types of agents as *flexible* agents and denote the fraction of the flexible agents by α_F. Since a flexible agent can serve both classes, he chooses the portion of his service capacity that is allocated to each class in addition to his pricing decision.

Once the firm chooses a pair of passing levels (ω_A, ω_B), the fraction of flexible and dedicated agents in marketplaces with a large number of agents can be approximated as follows:

$$\alpha_F(\omega_A, \omega_B) \simeq \int_{\omega_A}^{\bar{R}_A} ds_A \int_{\omega_B}^{\bar{R}_B} f_{A,B}(s_A, s_B) ds_B \ \textit{and} \ \alpha_i(\omega_A, \omega_B) \simeq \int_{\omega_i}^{\bar{R}_i} ds_i \int_{0}^{\omega_j} f_{A,B}(s_i, s_j) ds_j \qquad (1)$$

for all $i, j \in \{A, B\}$ with $j \neq i$. Throughout the paper, we use the term *skill-mix structure* to denote the portfolio of different agent types in the marketplace. For instance, a marketplace may consist of all three types of agents: the flexible agents, and the dedicated agents for each class. We refer to such a skill-mix structure as "*M-Network*." Moreover, there may be only flexible agents in a marketplace. This skill-mix structure will be referred to as "*V-Network*." In addition to these two, there may be other skill-mix structures such as "*N-Network*" and "*I-Network*." We illustrate these different structures in Figure 1. The moderating firm can set up various skill-mix structures by changing the passing levels ω_A and ω_B.

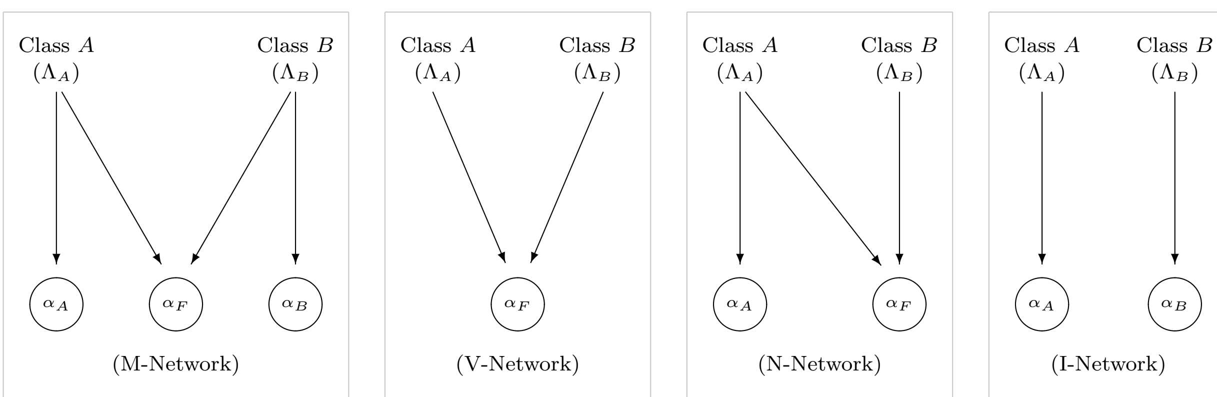


Figure 1 Different skill-mix structures that can be set up by the moderating firm.

In addition to changing the skill-mix structure, passing levels impact the expected reward that a customer earns upon her service completion. For example, when both passing levels are set to zero, a class $i \in \{A,B\}$ customer expects to earn the average value that agents generate for her class, which can be approximated by $\mathbf{E}[S_i]$. However, if passing levels are positive, customers can update their expected reward by knowing that the skills of eligible agents are above certain thresholds. More specifically, for any passing levels (ω_A, ω_B) and $i,j \in \{A,B\}$ with $j \neq i$, the expected reward that a class i customer earns from a dedicated agent becomes $\mathbf{E}[S_i|S_i \geq \omega_i, S_j < \omega_j]$ and can be written as follows:

$$R_i(\omega_A,\omega_B) = \Big(\int_{\omega_i}^{\bar{R}_i} ds_i \int_0^{\omega_j} s_i f_{A,B}(s_i,s_j)ds_j\Big)/\alpha_i(\omega_A,\omega_B) \tag{2}$$

Likewise, a class i customer expects to earn $R_{iF}(\omega_A,\omega_B)$ from a flexible agent given passing levels (ω_A,ω_B), where

$$R_{iF}(\omega_A,\omega_B) = \Big(\int_{\omega_i}^{\bar{R}_i} ds_i \int_{\omega_j}^{\bar{R}_j} s_i f_{A,B}(s_i,s_j)ds_j\Big)/\alpha_F(\omega_A,\omega_B). \tag{3}$$

Notice that in the above expressions, we allow customers to updated their expected rewards even if the firm does not offer the exam screening the skill they demanded. For instance, the expected rewards for class A customers from a dedicated and a flexible agent depend on the passing level in Exam B even if the firm does not offer Exam A.

In our model, we suppose that the moderating firm does not allow agents to serve customers when they fail a test for analytical tractability. In real service marketplaces, moderating firms may choose less severe actions to handle agents who fails in the tests. For example, upwork.com lets providers hide their test results and does not ban providers from serving customers when they fail a test. However, at the same time the default practice of the web-site is to publicize the test results when agents pass the exams. In other words, the web-site seems to have a preference for announcing the results of successful attempts and letting customers to differentiate between passing and failing agents. It is also important to note that, upwork.com offers guidelines to the service providers at Upwork Help Center (2016) where one of the first recommendations is "... take tests to demonstrate your skills to potential clients and make yourself more marketable ... for maximum impact, you may want to take tests corresponding with the skills you want to present to potential clients and considering hiding those where you rank below average ..." The Help Center also clearly warns the service providers about the possibility of not being hired if they do not score well on the tests related to the skills desired by the customers. We believe these guidelines and the default announcements of the passing scores indicate that the web-site indirectly tries to keep the failing agents outside of the marketplace and thus provide a justification for our assumption.

4.2. Matching Demand and Supply

In addition to setting up the skill-mix, the moderating firm provides a mechanism that efficiently matches customers and agents. This mechanism aims at reducing inefficiency due to the possibility that an individual customer may not find an idle agent on her own while there are available agents who can serve her. For instance, upwork.com achieves this goal by allowing customers to post their needs and allowing service providers to apply to these postings. When a customer posts a job at upwork.com, agents that are willing to serve this customer apply to the posting. Among the available applicants, the customers favor agents charging the lowest price. When there are not any immediate applications, the customers wait for agents to apply. The main driver of the efficiency in this setting is the fact that customers no longer need to specify an agent upon their arrival. The job posting mechanism allows customers to postpone their service request decisions until they have enough information about the availability of the providers.

Note that the expected utility of a customer will depend on both the price and the type of the agent who serves her because each agent type may provide a different expected reward. To account for that, we define the "*net reward*" of a class $i \in \{A, B\}$ customer from a dedicated agent and a flexible agent charging p as $R_i(\omega_A, \omega_B) - p$ and $R_{iF}(\omega_A, \omega_B) - p$ for any given pair of thresholds (ω_A, ω_B), respectively. Then, we model the efficiency improvement in the system by considering the marketplace as a multi-server queueing system where customers wait in a common queue and are matched with the agent offering the highest net-reward when there are available agents. When there are multiple agents offering the highest net-reward, customers are assigned to these agents randomly. In such a marketplace, the specifics of the customer decision making and experience will be as follows: An arriving class $i \in \{A, B\}$ customer first chooses whether to request service or not by observing the agents' pricing and service decisions. We denote the fraction of class $i \in \{A, B\}$ customers requesting service by D_i. If there are any available agents when the customer arrives, her service starts immediately and she obtains the highest net-reward offered by the available agents. Otherwise, the customer enters a queue and wait until an agent becomes available for her.

As we mentioned before, dedicated agents make only pricing decisions whereas flexible agents make both pricing and service decisions. We summarize pricing and service strategies of the agents by the vectors $(\mathbf{r_A}, \mathbf{y_A}, \mathbf{t_A}) \equiv (r_{A_n}, y_{A_n}, t_{A_n})_{n=1}^{N_A}$ and $(\mathbf{r_B}, \mathbf{y_B}, \mathbf{t_B}) \equiv (r_{B_n}, y_{B_n}, t_{B_n})_{n=1}^{N_B}$. We refer to the agents who offer the net reward r_{i_n} and allocate t_{i_n} portion of their capacities to class $i \in \{A, B\}$ as sub-pool i_n. Notice that t_{i_n} may be different than 1 only in sub-pools consisting of flexible agents. We let $y_{i_n}k$ be the number of agents in sub-pool i_n and N_i is the number of different sub-pools serving class $i \in \{A, B\}$. We also denote the fraction of total service capacity available for class $i \in \{A, B\}$ by $\overline{\alpha}_i \equiv \sum_{n=1}^{N_i} t_{in} y_{in}$. We suppose $r_{i_1} \geq \cdots \geq r_{i_{N_i}}$, without loss of generality, and assume that the agents cannot offer a net reward less than the outside option, i.e., $r_{i_{N_i}} \geq \underline{u}$ for any $i \in \{A, B\}$.[1] We illustrate our marketplace model in the following figure.

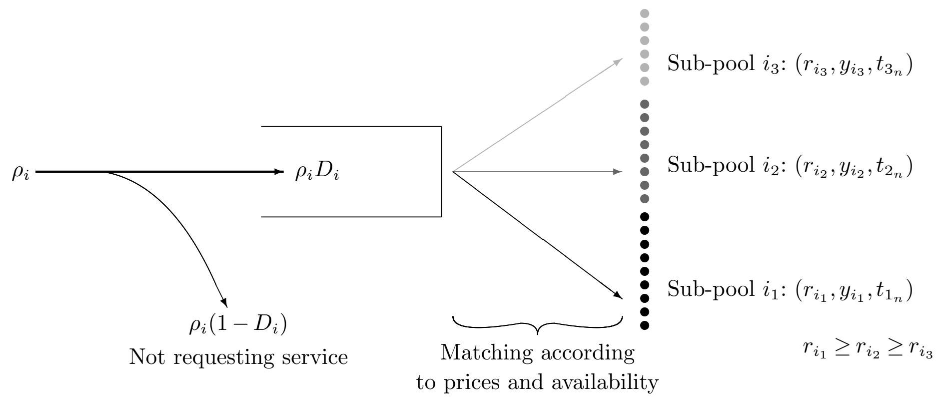


Figure 2 Illustration of the operations of the marketplace for class $i \in \{A, B\}$.

Under the model described above, the system that each class customers face operates like a generic $M/M/s$ system with the arrival rate of $s\rho$ and the service rate of 1 where ρ depends on the decisions of the customers and the agents. However, unlike a regular multi-server system, customers are assigned to the available agents offering the highest net reward based on the vector $(\mathbf{r}, \mathbf{y}) \equiv (r_n, y_n)_{n=1}^{N}$ where y_n is the fraction of agents offering the net reward r_n and N is the number of different sub-pools of agents. In this queueing system, we denote the expected waiting time by $W(\rho)$ as it does not depend on the agents' pricing decisions. We also let $P_\ell(\mathbf{r}, \mathbf{y}, \rho)$ be the probability with which a customer is served by an agent in sub-pool $\ell \in \{1, \ldots, N\}$ and $\sigma_\ell(\mathbf{r}, \mathbf{y}, \rho)$ be the utilization of agents in sub-pool ℓ. Then, we can write the expected utility of a class $i \in \{A, B\}$ customer requesting service as

$$\sum_{n=1}^{N_i} P_n(\mathbf{r_i}, \mathbf{t_i} \circ \mathbf{y_i}/\overline{\alpha}_i, \rho_i D_i/\overline{\alpha}_i) r_{i_n} - c_i W(\rho_i D_i/\overline{\alpha}_i), \tag{4}$$

as a function of the decisions of the customers and the agents, where the $\circ$ operator takes the element-wise product of two vectors. We can also write the revenue of a dedicated agent in sub-pool i_ℓ for any given passing levels (ω_A, ω_B) as

$$\left[R_i(\omega_A, \omega_B) - r_{i_\ell}\right] \sigma_\ell(\mathbf{r_i}, \mathbf{t_i} \circ \mathbf{y_i}/\overline{\alpha}_i, \rho_i D_i/\overline{\alpha}_i) \tau_i, \tag{5}$$

[1] One may consider a model where agents are allowed to offer less than $\underline{u}$. We can show that such a pricing strategy cannot emerge as an equilibrium. Thus, our key findings would continue to hold, despite additional analytical cumbersome.

Similarly, the revenue of a flexible agent in the sub-pool i_ℓ is

$$t_{i_\ell}\left[R_{iF}(\omega_A,\omega_B)-r_{i_\ell}\right]\sigma_\ell(\mathbf{r_i},\mathbf{t_i}\circ\mathbf{y_i}/\overline{\alpha}_i,\rho_i D_i/\overline{\alpha}_i)\tau_i. \tag{6}$$

In the above expressions, we adjust the fraction of agents in sub-pool i_ℓ as $t_{i_\ell}y_{i_\ell}/\overline{\alpha}_i$ and the demand rate as $\rho_i D_i/\overline{\alpha}_i$ to fit our original model to the generic $M/M/s$ we described above.

After describing the operational tool provided by the moderating firm, we now discuss the moderating firm's strategic tool, which changes the nature of the competition among agents. In a marketplace such as upwork.com, service providers are offered discussion boards where they are allowed to exchange information. Moreover, the firm supports the creation of affiliation groups, which are self-enforcing entities. Motivated by these examples, we assume that the moderating firm allows agents to make non-binding communication prior to making their decisions, so that the players can discuss their strategies but are not allowed to make binding commitments. The economics literature suggests that the stability of any equilibrium outcome can be threatened by potential deviations formed by coalitions, even in noncooperative games, due to pre-play communications (See Bernheim et al. (1987), Ray (1996) and Moreno and Wooders (1996)). In other words, players can try to self-coordinate their actions in a mutually beneficial way when the moderating firm allows them to communicate among themselves, despite the fact that each agent selfishly maximizes his own utility. As it is discussed in Allon et al. (2012), this can be modeled by an equilibrium concept which allows several agents to deviate together instead of deviating individually. However, self-coordination of agents is restricted because the marketplaces we consider tend to be large, and thus, the sizes of group deviations are limited. We denote the largest fraction of agents that can deviate together by $\delta \leq 1$.

In this section, we outline the different roles of the moderating firm and discuss how these roles affect the structure of the marketplace. Next, we model the strategic interaction between the agents and the customers as a sequential-move game given the setup of Section 3, along with the above mentioned roles of the moderating firm. We also characterize the equilibrium outcome of a special marketplace structure.

5. The Game Between Customers and Agents

In this section, we formally set up the two-stage game between the agents and the customers based on the model introduced in Section 3 and the roles of the moderating firm mentioned in Section 4. As the first step of the strategic interaction between agents and customers, agents make and announce their service and pricing decisions. Then, in the second stage, each arriving customer observes these decisions and decides whether to request service or not. We suppose that agents make the first move because agents have permanent profile pages, where they post pricing information for their services, in real online marketplaces. Moreover, posting a price is neither a requirement nor a binding decision for customers in job postings. Therefore, customers usually do not announce any price in job postings because they prefer to wait and see the agents' availability and prices. We refer to the equilibrium among customers in the second stage as *Customer Equilibrium* and the equilibrium of the whole game as *Market Equilibrium*.

In analyzing the equilibrium outcome of marketplaces with finite number of agents, we would like to highlight the following two observations: 1) The arising system dynamic is too complex for the exact analysis. We need to keep track of the number of busy agents in each sub-pool in order to obtain agent utilizations, and this requires us to have a multi dimensional state space. 2) Asymptotic analysis is applicable since these marketplaces tends to be large. Thus, in this paper, we shall approximate the original system by a fluid model where the number of agents k goes to infinity while the demand-supply ratios are remaining constant. The benefit of using a fluid model is that it provides an accurate yet simple approximation for the expected waiting time and utilization functions and thus helps us to derive the utility of the customers and agents in simple form. We build our fluid model based on the framework introduced in Whitt (2006). Whitt (2006) shows that the expected waiting time function $W(\rho)$ we describe in Section 4.2 can be approximated as:

$$W(\rho) \simeq W^f(\rho) \equiv \begin{cases} 0 & if\ \rho \leq 1 \\ \infty & if\ \rho > 1. \end{cases}$$

Whitt (2006) also provides approximations for the utilization of agents, which is simply equal to $\min\{\rho, 1\}$ in a system with the system load of ρ. This result implies that servers can be fully utilized only when the system is overloaded, i.e., $\rho > 1$, otherwise agents will be underutilized and their utilization will be equal to the demand-supply ratio. Unfortunately, we cannot use this result directly to obtain agent utilizations. In our model, customers are matched with the available agents who offer the highest net-reward whereas Whitt (2006) studies a queueing system where the customers are assigned to available agents randomly. Therefore, we propose a modified version of the approximation in Whitt (2006) for the agent utilizations. Namely, we assume that in the fluid approximation of the generic $M/M/s$ system we describe in Section 4.2, agents in sub-pool $\ell \in \{1, \ldots, N\}$ will be able to serve some customers only if the arrival rate is greater the service capacity available by the sub-pools offering a net-reward higher than sub-pool ℓ does, which occurs when $\rho > \sum_{n=1}^{\underline{\ell}-1} y_n$, with $\underline{\ell} \equiv \min\{n : r_n = r_\ell\}$. Otherwise, the utilization of sub-pool ℓ will be zero because there will always be enough available agents offering a net-reward higher than sub-pool

ℓ, and thus customers will be matched with those agents. Once ρ exceeds $\sum_{n=1}^{\underline{\ell}-1} y_n$, the rate of customers served by sub-pool ℓ will be the maximum of the demand left from the sub-pools offering a higher net-reward, which is $\rho - \sum_{n=1}^{\underline{\ell}-1} y_n$, and the service capacity of sub-pool ℓ and other sub-pools offering the same net reward, which is $\sum_{n=\underline{\ell}}^{\bar{\ell}} y_n$, with $\bar{\ell} \equiv \max\{n : r_n = r_\ell\}$. Hence, the utilization of the agents in sub-pool ℓ can be approximated as follows:

$$\sigma_\ell(\mathbf{r},\mathbf{y},\rho) \simeq \sigma_\ell^f(\mathbf{r},\mathbf{y},\rho) \equiv \begin{cases} 0 & \textit{if } \rho \leq \sum_{n=1}^{\underline{\ell}-1} y_n \\ \frac{\rho - \sum_{n=1}^{\underline{\ell}-1} y_n}{\sum_{n=\underline{\ell}}^{\bar{\ell}} y_n} & \textit{if } \sum_{n=1}^{\underline{\ell}-1} y_n < \rho < \sum_{n=1}^{\bar{\ell}} y_n \\ 1 & \textit{if } \rho \geq \sum_{n=1}^{\bar{\ell}} y_n. \end{cases}$$

In the fluid model, we also have that the probability of being served by sub-pool ℓ becomes the rate of the customer demand served by this sub-pool, which is $y_\ell \sigma_\ell^f(\mathbf{r},\mathbf{y},\rho)$, to total rate of customers being served, which is the minimum of the arrival rate ρ and total service capacity, which is normalized to 1 in the generic $M/M/s$ system we consider. Hence, using the approximation for the agent utilizations, we can approximate $P_\ell(\mathbf{r},\mathbf{y},\rho)$ function as

$$P_\ell(\mathbf{r},\mathbf{y},\rho) \simeq P_\ell^f(\mathbf{r},\mathbf{y},\rho) \equiv \frac{y_\ell \sigma_\ell^f(\mathbf{r},\mathbf{y},\rho)}{\min\{\rho, 1\}}.$$

Based on the above approximations, we can, then, approximate the expected utility of a class $i \in \{A,B\}$ customer requesting service in our original model as:

$$U_i^f(\mathbf{r_i},\mathbf{y_i},\mathbf{t_i},D_i) = \sum_{n=1}^{N_i} P_n^f(\mathbf{r_i},\mathbf{t_i}\circ\mathbf{y_i}/\overline{\alpha}_i, \rho_i D_i/\overline{\alpha}_i) r_{i_n} - c_i W^f(\rho_i D_i/\overline{\alpha}_i), \tag{7}$$

given the service decisions of class $i \in \{A,B\}$ customers, D_i, and the strategies of agents serving class i, which is $(\mathbf{r_i},\mathbf{y_i},\mathbf{t_i})$. Similarly, for any given passing levels (ω_A,ω_B), the revenue of a dedicated and a flexible agent in sub-pool i_ℓ can be approximated as

$$\begin{aligned} &\left[R_i(\omega_A,\omega_B) - r_{i_\ell}\right]\sigma_\ell^f(\mathbf{r_i},\mathbf{t_i}\circ\mathbf{y_i}/\overline{\alpha}_i, \rho_i D_i/\overline{\alpha}_i)\tau_i, \\ \textit{and} \quad &\left[R_{iF}(\omega_A,\omega_B) - r_{i_\ell}\right]\sigma_\ell^f(\mathbf{r_i},\mathbf{t_i}\circ\mathbf{y_i}/\overline{\alpha}_i, \rho_i D_i/\overline{\alpha}_i)\tau_i, \ \textit{respectively.} \end{aligned}$$

Once we introduce the fluid approximation of the marketplace, we next formally describe and study the *Customer Equilibrium* and the *Market Equilibrium* in the fluid model. Since we consider a two-stage game, we start with the equilibrium among customers given the strategy of agents.

5.1. Customer Equilibrium

As we mentioned before, customers make their service request in order to maximize their expected utility. Therefore, a customer from class $i \in \{A,B\}$ requests service if her expected utility from joining the marketplace (weakly) exceeds her outside utility of $\underline{u}$. The first condition of the *Customer Equilibrium* captures this requirement as we formally define as follows:

DEFINITION 1 (CUSTOMERS EQUILIBRIUM). Given any $(\mathbf{r_i}, \mathbf{y_i}, \mathbf{t_i})$ for $i \in \{A, B\}$, we say that D_i is a *Customers Equilibrium* if

$$D_i = \arg\max\{D : U_i^f(\mathbf{r_i}, \mathbf{y_i}, \mathbf{t_i}, D_i) \geq \underline{u}, D \leq 1\}.$$

In other words, in a *Customers Equilibrium*, class i customers keep requesting service as long as doing so is not strictly worse than their outside option.

In addition to stating that customers' expected utility must be at least $\underline{u}$, the *Customer Equilibrium* ensures that customers join the system unless any additional demand makes the expected utility strictly less than the outside option. This feature essentially breaks the tie for a customer who is indifferent between joining the system and leaving immediately in favor of joining, and thus ensures the uniqueness of *Customer Equilibrium* given the pricing decisions of the agents. We formally present the *Customer Equilibrium* in Proposition 1.

PROPOSITION 1. *Given any* $(\mathbf{r_i}, \mathbf{y_i}, \mathbf{t_i})$*, let* D_i^{ce} *be a Customer Equilibrium for class* $i \in \{A, B\}$*. Then, we have that*

$$D_i^{ce} = \begin{cases} 1 & \text{if } \rho_i \leq \overline{\alpha}_i, \\ \overline{\alpha}_i/\rho_i & \text{if } \rho_i > \overline{\alpha}_i. \end{cases}$$

Furthermore, letting $\sigma_{i_\ell}^{ce}(\mathbf{r_i}, \mathbf{y_i}, \mathbf{t_i})$ *be the utilization of agents in sub-pool* i_ℓ*, with* $\ell \in \{1, \dots, N_i\}$*, in a Customer Equilibrium, we have that*

$$\sigma_{i_\ell}^{ce}(\mathbf{r_i}, \mathbf{y_i}, \mathbf{t_i}) = \begin{cases} 0 & \text{if } \rho_i \leq \sum_{n=1}^{\underline{\ell}_i - 1} t_{i_n} y_{i_n} \\ \frac{\rho_i - \sum_{n=1}^{\underline{\ell}_i - 1} t_{i_n} y_{i_n}}{\sum_{n=\underline{\ell}_i}^{\bar{\ell}_i} t_{i_n} y_{i_n}} & \text{if } \sum_{n=1}^{\underline{\ell}_i - 1} t_{i_n} y_{i_n} < \rho_i < \sum_{n=1}^{\bar{\ell}_i} t_{i_n} y_{i_n}, \\ 1 & \text{if } \rho_i \geq \sum_{n=1}^{\bar{\ell}_i} t_{i_n} y_{i_n}, \end{cases}$$

where $\underline{\ell}_i \equiv \min\{n : r_{i_n} = r_{i_\ell}\}$ *and* $\bar{\ell}_i \equiv \max\{n : r_{i_n} = r_{i_\ell}\}$*.*

The above proposition shows that the *Customer Equilibrium* depends on whether the arrival rate form class $i \in \{A, B\}$, which is $\rho_i \tau_i$, is greater than the service capacity available for class, which is $\overline{\alpha}_i \tau_i$. For any $\rho_i \leq \overline{\alpha}_i$, all class i customers request service because they will obtain a net reward that (weakly) exceeds their outside option even if all of them join the marketplace. On the other hand, some of the class i customers have to stay outside the marketplace when $\rho_i > \overline{\alpha}_i$.

Proposition 1 also establishes the utilization of the agents serving class $i \in \{A, B\}$ in a *Customer Equilibrium*. We show that the utilization of a sub-pool can be non-zero only if the arrival rate from class $i \in A, B$ is greater than the total service capacity of the sub-pools offering a net reward higher than the sub-pool offers. Furthermore, a sub-pool can be fully utilized when $\rho_i \tau$ exceeds the total service capacity of the sub-pools offering a higher net reward by at least its capacity. It is worth noting that the agent utilizations do not depend on the equilibrium decisions of the customers. The intuition behind this result is that the agent utilizations may depend on the customers' decisions only if some of the customers do not request service, but the rate of class $i \in \{A, B\}$ customers requesting service in the equilibrium is always equal to the total service capacity available for class i when customers ration themselves. This can be seen from the fact that $\rho_i D_i^{ce} = \overline{\alpha}_i$.

Once we characterize the equilibrium among customers and obtain the revenues of agents in this equilibrium, we now focus on the first stage of the game.

5.2. Market Equilibrium

The *Customer Equilibrium* we study in the previous subsection lets us derive the agent revenues when we fix the service and pricing decisions of the agents. Using this result, we now study the equilibrium outcome of the whole game, which will be referred to as the *Market Equilibrium.* To this end, we need to find the service and pricing decisions from which agents have no incentive to deviate in the first stage. Agents can deviate by either joining an existing sub-pool or announcing a new price. Furthermore, a limited fraction of agents is allowed to deviate together since the moderating firm enables communication among agents. Therefore, an equilibrium in the first stage should be immune to any of these two types of deviations formed by at most δ fraction of agents.

In the large-scale marketplaces we study, it is possible that a small group of agents can find profitable deviation from every price in some cases. However, these deviations require infinitesimally small price changes, which might be unrealistic in real service marketplaces. Thus, we ignore such deviations by restricting the set of prices agents can charge. To be more specific, we suppose that the agents must choose their prices from a finite set where price increments are ϵ, a small number close to zero. Then, we focus on an equilibrium concept which requires immunity only against deviations within this finite price sets as formally stated in Definition 2. To ease notation, we denote $R_i(\omega_A, \omega_B)$ and $R_{iF}(\omega_A, \omega_B)$ by R_i and R_{iF}, respectively, for any given passing levels (ω_A, ω_B) and any $i \in \{A, B\}$.

Definition 2 (Market Equilibrium). Let $(\mathbf{r_i}, \mathbf{y_i}, \mathbf{t_i}) \equiv (r_{i_n}, y_{i_n})_{n=1}^{N_i}$ summarize the strategy of all agents in the marketplace for any $i \in \{A, B\}$. Then, $(\mathbf{r_i}, \mathbf{y_i}, \mathbf{t_i})$ is a (ϵ, δ)-*Market Equilibrium* (*(ϵ, δ)-ME*) if the following conditions are satisfied.

1. For any $\ell \leq N_i$, $0 < d \leq \min\{y_{i_\ell}, \delta\}$, $i \in \{A, B\}$, and $r' \in \{\underline{u}, \underline{u} + \epsilon, \underline{u} + 2\epsilon, \dots\} \setminus r_{i_\ell}$,

$$[R_i - r_{i_\ell}]\sigma_{i_\ell}^{ce}(\mathbf{r_i}, \mathbf{y_i}, \mathbf{t_i}) \geq [R_i - r']\sigma_{i_{\ell'}}^{ce}(\mathbf{r'_i}, \mathbf{y'_i}, \mathbf{t'_i}), \tag{8}$$

$$[R_{iF} - r_{i_\ell}]\sigma_{i_\ell}^{ce}(\mathbf{r_i}, \mathbf{y_i}, \mathbf{t_i}) \geq [R_{iF} - r']\sigma_{i_{\ell'}}^{ce}(\mathbf{r'_i}, \mathbf{y'_i}, \mathbf{t'_i}), \tag{9}$$

where $(\mathbf{r_i'}, \mathbf{y_i'}, \mathbf{t_i'})$ is the strategy of all agents serving class i and $i_{\ell'}$ is the new sub-pool of deviating agents after the deviation. In other words, any small group of agents from any sub-pool cannot improve their revenues by changing their prices.

2. For any $\ell \leq N_i$, $0 < d \leq \min\{y_{i_\ell}, \delta\}$, $0 < t \leq t_{i_\ell}$, $i \in \{A, B\}$, $j \in \{A, B\} \setminus i$, and $r' \in \{\underline{u}, \underline{u} + \epsilon, \underline{u} + 2\epsilon, \dots\}$,

$$\begin{aligned} t_{i_\ell}[R_{iF} - r_{i_\ell}]\sigma^{ce}_{i_\ell}(\mathbf{r_i}, \mathbf{y_i}, \mathbf{t_i})\tau_i \geq\ & (t_{i_\ell} - t)[R_{iF} - r_{i_\ell}]\sigma^{ce}_{i_\ell}(\mathbf{r_i'}, \mathbf{y_i'}, \mathbf{t_i'})\tau_i \\ & + t[R_{jF} - r']\sigma^{ce}_{j_{\ell'}}(\mathbf{r_j'}, \mathbf{y_j'}, \mathbf{t_j'})\tau_j, \end{aligned} \tag{10}$$

where $(\mathbf{r_i'}, \mathbf{y_i'}, \mathbf{t_i'})$ and $(\mathbf{r_j'}, \mathbf{y_j'}, \mathbf{t_j'})$ are the strategy of all agents and $j_{\ell'}$ is the new sub-pool of deviating agents after the deviation. In other words, any small group of flexible agents cannot improve their revenues by changing their service decisions.

Moreover, $(\mathbf{r_i}, \mathbf{y_i}, \mathbf{t_i})$ is a *Market Equilibrium* if there exists a sequence $(\mathbf{r_i}^k, \mathbf{y_i}^k)$ such that $(\mathbf{r_i}^k, \mathbf{y_i}^k)$ is a (ϵ^k, δ^k)-*ME* where $\epsilon^k \to 0$, $\delta^k \to 0$, and $(\mathbf{r_i}^k, \mathbf{y_i}^k) \to (\mathbf{r_i}, \mathbf{y_i}, \mathbf{t_i})$ for all $n \leq N_i$ as $k \to \infty$.

(8) in the (ϵ, δ)-*ME* definition states that dedicated agents have no incentive to deviate. Note that dedicated agents cannot change the customer class they serve. Therefore, (ϵ, δ)-*ME* accounts for two possible deviations for dedicated agents: joining an existing sub-pool or creating a new one. On the other hand, the flexible agents have the option of choosing the portion of service capacity they allocate for each class. Thus, (ϵ, δ)-*ME* ensures that the flexible agents cannot improve their revenues whether they change how they distribute their capacity among customer classes. In particular, (9) focuses on flexible agent deviations when they keep their service decisions unchanged, whereas (10) considers deviations where the flexible agents change how much service capacity they allocate to each class. Finally, we conclude that a strategy profile is a *Market Equilibrium* if it is the limit of a sequence of (ϵ, δ)-*ME* as ϵ and δ become arbitrarily small. It is important to note that our main equilibrium concept is *Market Equilibrium*. By considering *Market Equilibrium* as the limit of a sequence of (ϵ, δ)-*ME*, we lessen the role of our two previous assumptions: i) the agents must choose their prices from a finite set with a price increment of ϵ, and ii) δ fraction of agents can deviate together.

It is quite tedious to study the *Market Equilibrium* of the whole marketplace in detail because the moderating firm can create many different *skill-mix structures* as discussed in Section 4.1. Hence, we next characterize the equilibrium outcome in a marketplace with a simplified market structure, namely one customer class and two types of agents. This structure constitutes the building block of a marketplace with two classes. Carrying out our analysis in this building block model is a fundamental step towards finding the equilibrium outcome of the whole marketplace. It also allows us to discuss the intuition behind our results in a more transparent way.

5.3. A Marketplace with One Class of Customers

In this subsection, we focus on a simplified version of our marketplace model. To be specific, we consider a marketplace where there is only one class of customers, but two types of agents, say high- and low-value. There are $\alpha_H k$ high-value and $\alpha_L k$ low-value agents, and the service rate of all agents are τ_s. We assume that the arrival rate of customers is Λ_s and denote the ratio $\Lambda_s/(\tau_s)k$ by ρ_s. Furthermore, customers earn a reward of R_H when their service is completed by a high-value agent, and they earn R_L when a low-value agent serves them. We suppose $R_H \geq R_L$ for ease of explanation. Our results would only need relabeling if $R_L \geq R_H$. Finally, we denote the waiting cost by c_s. We keep all other assumptions we made in Section 3. We use the equilibrium concepts introduced in Sections 5.1 and 5.2 and establish the equilibrium revenues of the agents in the fluid model. We refer to the marketplace as *buyer's market* if $\rho_s < \alpha_H + \alpha_L$ and *seller's market* otherwise.

As we mentioned before, the moderating firm may set up the skill-mix structure in the marketplace (i.e., the capacity of the dedicated and the flexible agents) by changing the passing levels in each exam. Figure 1 illustrates the possible skill-mix structures that can arise based on the firm's skill screening decision. One can use the aforementioned simplified marketplace model to study and derive the equilibrium outcomes in any of these skill-mix structure. For instance, once the agents make their service decisions in an *M-Network*, we can study the part of the marketplace related to each class in isolation as illustrated in Figure 3. Therefore, as a first step toward characterizing the equilibrium of the entire marketplace, we study the firm's business with each customer class separately.

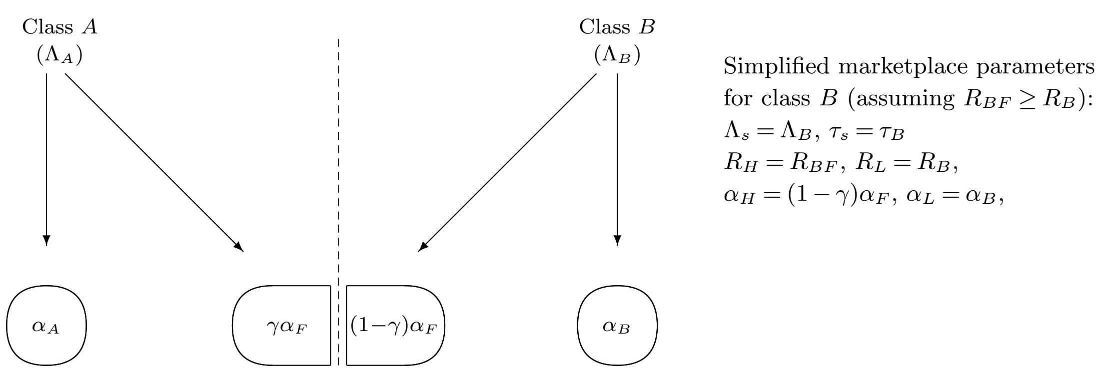


Figure 3 **Illustration of using the simplified market model to study each customer class in isolation given that flexible agents allocate γ fraction of their capacity to class A.**

Theorem 1 formally presents the *Market Equilibrium* in the simplified market model. An important implication of our equilibrium result is that the revenue of the low-value agents cannot exceed their operating cost, which is normalized to zero, in a buyer's market. The main driver of this result is that the high-value agents can price the low-value agents out of the market when $\rho_s < \alpha_H + \alpha_L$ since customers earn a higher reward from the high-value agents. It is also worth noting that the low equilibrium revenues of the low-value agents create a revenue cap for the high-value agents in a buyer's market. On the other hand, both groups of agents can agree to charge the customer reward they generate minus the outside option in a seller's market. In such an equilibrium, agents charge their highest prices and are fully utilized, and the rate of customers requesting service is equal to the total capacity, $\alpha_H + \alpha_L$. In other words, the equilibrium behavior of the agents ensures that the demand perfectly matches supply in the system and the agents can extract all of the customer surplus. The intuition behind these high prices is the following: Even if agents charge lower than the highest price that they can charge, the customer demand in equilibrium cannot exceed the total service capacity because the waiting times would explode otherwise. As a result, customers leave the system with strictly positive surplus. This allows a small group of agents to increase their prices and improve their revenues as long as the price increase.

Theorem 1. *Let V_H^{sm} and V_L^{sm} be the revenue of a high-value and a low-value agent, respectively, in a Market Equilibrium of a marketplace with one customer class and two agent pools.*

1. *If $\rho_s < \alpha_H$, then we have that $V_H^{sm} = V_L^{sm} = 0$.*
2. *If $\rho_s = \alpha_H$, then we have that $V_H^{sm} \leq (R_H - R_L)\tau$, and $V_L^{sm} = 0$.*
3. *If $\alpha_H < \rho_s < \alpha_H + \alpha_L$, then we have that $V_H^{sm} = (R_H - R_L)\tau$, and $V_L^{sm} = 0$.*
4. *If $\rho_s = \alpha_H + \alpha_L$, then we have that $V_H^{sm} = V_L^{sm} + (R_H - R_L)\tau$, and $V_L^{sm} \leq (R_L - \underline{u})\tau$.*
5. *If $\rho_s > \alpha_H + \alpha_L$, then we have that $V_H^{sm} = (R_H - \underline{u})\tau$, and $V_L^{sm} = (R_L - \underline{u})\tau$.*

The marketplace we consider in this subsection is similar to the one with nonidentical agents studied in Allon et al. (2012). Allon et al. (2012) develops an asymptotic theory to understand the behavior of the equilibrium along the sequence of marketplaces growing in size. The methodology developed in Allon et al. (2012) can easily become analytically intractable while analyzing more complex problems such as the moderating firm's skill management problem. Thus in this paper, we focus directly on the limiting game, whose results are easy to incorporate into the moderating firm's problem. One can see that our findings in Theorem 1 are aligned with the results in Allon et al. (2012), which provides a strong support that our fluid approximation captures the main managerial insights obtained from an asymptotic analysis. Furthermore, given that we use a fluid model, we also investigate the outcomes of asymmetric equilibria, which are ignored in Allon et al. (2012), in Appendix S.1 of the online supplement. We show that any asymmetric *Market Equilibrium* is outcome-equivalent (from the agents' point of view) to a symmetric *Market Equilibrium* where the same type of agents charge their operating costs, which is normalized to zero.

After characterizing the equilibrium outcome in our building block model, we next turn our attention to the moderating firm's skill-mix and capacity decisions.

6. The Moderating Firm's Problem

In the previous section, we analyze a model where the skill-mix of the marketplace is given. In this section, we study the firm's problem of finding the best skill-mix structure that maximizes its revenue. The firm's revenue is a predetermined share of the total revenue in the marketplace. Thus, the firm has to maximize the total revenue in the marketplace by choosing the appropriate passing levels (ω_A, ω_B). Once the firm chooses a skill-mix structure via the skill tests, the flexible agents make their service decisions. Based on the service decisions of the flexible agents, the firm's business with each class can be considered in isolation. In other words, we can view the firm as managing two marketplaces each of which follows the simplified market structure we study in Section 5.3. Then, we can calculate the revenue generated in the marketplace using Theorem 1. Letting $\overline{V}_F(\omega_A, \omega_B)$ be the equilibrium revenues of a flexible agent and $\overline{V}_{iD}(\omega_A, \omega_B)$ be the equilibrium revenues of a dedicated agent serving class $i \in \{A, B\}$, the total revenue of the marketplace for any passing levels (ω_A, ω_B) is[2]

$$\Pi(\omega_A, \omega_B) = k\big[\alpha_F(\omega_A, \omega_B)\overline{V}_F(\omega_A, \omega_B) + \sum_{i\in\{A,B\}} \alpha_i(\omega_A, \omega_B)\overline{V}_{iD}(\omega_A, \omega_B)\big].$$

The main focus of this paper is to gain a better understanding of the effectiveness of skill screening as a revenue management tool. As a first step toward this goal, we study the firm's problem under the following three cases: (i) *Benchmark*, where $\omega_A = \omega_B = 0$, (ii) *One-Test*, where either $\omega_A = 0$ or $\omega_B = 0$ but not both, and (iii) *Two-Tests*, where both $\omega_A > 0$ and $\omega_B > 0$. Considering these three cases separately allows us to analyze the revenue improvements in the marketplace after each additional test. We, then, find the optimal number of exams when the skill screening is costly.

A major technical challenge in finding the firm's optimal decisions is that the revenue function may have different functional forms in different regions of passing level space $[0, \bar{R}_A] \times [0, \bar{R}_B]$ because equilibrium revenues of the agents can change significantly even for slight adjustments in

[2] Theorem 1 shows that there might be multiple equilibria if the demand rate is equal to the capacity. For mathematical convenience, we focus on the equilibrium generating the highest possible revenue among all possible equilibria. The firm can sustain an equilibrium where the revenue arbitrarily close to the highest level among the multiple equilibria by perturbing the passing levels (ω_A, ω_B) slightly and creating a seller's market with a unique equilibrium. Moreover, Proposition 4 in Appendix S.2 shows that the flexible agents must earn the same revenue for any given equilibrium. Therefore, we can denote the revenue of the flexible agents by a unique value.

the passing levels. Finding the optimal passing levels, then, requires comparing all these different functional forms, which becomes analytically intractable when the skills, S_A and S_B, follow a general joint probability distribution. For tractability of the firm's problem, we need to impose some assumptions on the skill distributions.

The marketplaces we study, in general, attract service providers who are endowed with similar skills. For example, upwork.com has many agents with programming skills on web and mobile app development. Therefore, it is not unrealistic to suppose that agent skills are positively correlated. We also observe that the agent skills are mostly positively correlated from the data we collected from upwork.com. In Table 1, we present the correlation between a pair of exams that are taken by at least 500 unique service providers.

	1	2	3	4	5	6	7	8	9	10	11
1: U.S. Eng. Skills	1.00										
2: English Spelling	0.34	1.00									
3: Office Skills	0.57	0.29	1.00								
4: Email Etiquette	0.56	0.37	0.58	1.00							
5: HTML 4.01	0.40				1.00						
6: PHP5					0.49	1.00					
7: Customer Service	0.54	0.30	0.61	0.60			1.00				
8: CSS 2.0					0.56			1.00			
9: Search Engine Opt.		0.17							1.00		
10: Call Cent. Skills	0.49	0.38		0.61			0.54			1.00	
11: UK Eng. Skills	0.79	0.34									1.00

Table 1 Correlation between exam pairs that are taken at least 500 times

We also plot the frequency of the correlation coefficients between exam pairs that are taken together at least 100 times in Figure 4. As one can see from this figure, correlation coefficients between exam pairs range between 0.1 (almost independent) and 0.8 (almost perfectly correlated). In order to capture these observations, we focus on a family of joint skill distributions with a shape parameter η. Specifically, we suppose $f_{A,B}(s_A, s_B; \eta) = \frac{\eta+1}{\eta-1}$ for any $0 \leq S_A^{\eta} \leq S_B \leq S_A^{1/\eta} \leq 1$ and zero otherwise for a given shape parameter $1 < \eta < \infty$.

As we formally present in Proposition 2, the correlation coefficient between S_A and S_B, denoted by $Corr(S_A, S_B)$, varies between zero and one. Furthermore, agent skills become independent and identically distributed as η approaches ∞ and perfectly correlated as η approaches one. We also note that the marginal distributions of S_A and S_B are symmetric, so that the averages value that an agent's service generates for both classes are the same, i.e., $\mathbf{E}[S_A] = \mathbf{E}[S_B]$. Henceforth, we will refer to these average values as *average skill* and denote it by $\mathbf{E}[S_\eta]$ for any given shape parameter η. We also let $F_\eta(\cdot)$ denote the marginal distribution of each agent skill.

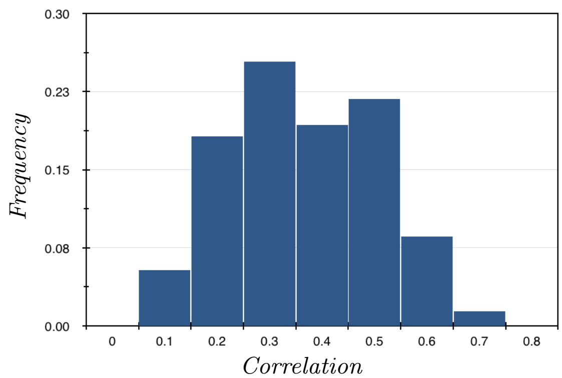


Figure 4 Frequencies of correlation coefficients in upwork.com data.

PROPOSITION 2. *For any* $1 < \eta < \infty$, *we have that* $0 < Corr(S_A, S_B) < 1$. *Furthermore, we have that*

i) $\lim_{\eta \to 1} Corr(S_A, S_B) = 1$,

ii) $\lim_{\eta \to \infty} Corr(S_A, S_B) = 0$, *and*

iii) $\lim_{\eta \to \infty} P(S_A \leq s_A, S_B \leq s_B) = \lim_{\eta \to \infty} P(S_A \leq s_A) P(S_B \leq s_B)$ *for any* $0 \leq s_A \leq 1$ *and* $0 \leq s_B \leq 1$.

In the following subsections, we restrict our attention to a market where the total demand rate exceeds the total service capacity of candidate agents, and normalize the utility of customers from their outside option, $\underline{u}$, to zero. We provide a brief discussion on the firm's optimal decisions when these assumptions are relaxed in the Conclusion. We also suppose $\rho_A \geq \rho_B$ without loss of generality. It is also worth mentioning that our methodology and model allow us to study negatively correlated skills. For instance, we could focus on the skills that satisfy $0 \leq 1 - S_A^{1/\eta} \leq S_B \leq 1 - S_A^{\eta} \leq 1$.

6.1. Benchmark

We start our analysis by considering the *Benchmark* case, where the firm does not offer any skill tests. One can view this case as if passing levels are set to zero, i.e. $(\omega_A, \omega_B) = (0, 0)$. We denote the total revenue of the marketplace in the *Benchmark* case by Π^o.

In the *Benchmark* case, all agents are eligible to serve both customer classes and can choose the customer class they would like to serve. In other words, the skill-mix structure is a *V-Network* with only flexible agents. Furthermore, customers' expected reward is equal to the average skill $\mathbf{E}[S_\eta]$ in the absence of any skill tests.

As we have only flexible agents in the *Benchmark* case, each class of customers face only one pool of agents after the agents make their service decisions. Thus, we can view the whole marketplace as two independent marketplaces once we have the decisions of the agents. Furthermore, both classes expect a reward of $\mathbf{E}[S_\eta]$ since the firm does not offer any test. Then, we can use Theorem 1 to determine the agent revenues in both of these marketplaces. To be specific, agents can extract all of the customer surplus by charging $\mathbf{E}[S_\eta]$ if the capacity they allocate to class $i \in \{A, B\}$ is less the demand from this class and earn zero otherwise. Hence, agents will always prefer to be in a seller's market, where the service capacity is scarce.

When the service rates for both classes are the same, i.e., when $\tau_A = \tau_B$, flexible agents can, indeed, sustain an equilibrium where customers from both classes face a seller's market due to the fact that the total demand rate is higher than potential service capacity, i.e., $\rho_A + \rho_B \geq 1$. Furthermore, in such an equilibrium, customers pay their expected reward and are left with no surplus. On the other hand, if $\tau_A \neq \tau_B$, an equilibrium where all flexible agents operate in a seller's market cannot emerge. In a seller's market, flexible agents would earn different equilibrium revenues from different classes because of the non-identical service rates, and this would contradict with our equilibrium concept, which requires that the flexible agents must earn the same revenue from each class when they serve both classes. Hence, when the service rates are different, flexible agents can extract the customer surplus fully only from one class, which turns out to be the class with the lower service rate. While serving the class with the highest service rate, the flexible agents have to charge a price that is less than $\mathbf{E}[S_\eta]$, the highest price they can charge. The following theorem formally presents these results and the total revenue of the marketplace under the *Benchmark* case.

THEOREM 2. *The total revenue of the marketplace in the Benchmark case is*

$$\Pi^o = \min\{\tau_A, \tau_B\} \frac{(\eta+1)^2}{(\eta+2)(2\eta+1)} k.$$

6.2. One Skill Test

After analyzing the *Benchmark* case, we now study the *One-Test* case, where the firm offers only one skill test. As each agent has two skills, the firm has to choose the skill to be tested and the passing level of the test. The firm's objective, in the *One-Test* case, is then to maximize the revenue, $\Pi(\omega_A, \omega_B)$, given the constraints of $\omega_A \omega_B = 0$ and $\omega_A + \omega_B > 0$.

In the *One-Test* case, agents who pass the skill test are eligible to serve both classes while the failing ones can only serve the customers who request service related to the skill that is not tested. In other words, the skill-mix structure is an *N-Network* with a flexible and a dedicated agent pool. Throughout this section, we suppose the firm offers only Exam $i \in \{A, B\}$, so that dedicated agents are only eligible to serve class $j \neq i \in \{A, B\}$ customers. We denote the optimal revenue of the marketplace by Π_i^* and the optimal passing level by ω_i^* . We also define the relative improvement in revenue from the *Benchmark* case to the *One-Test* case as $\Pi_i^*/\Pi^o - 1$ and denote it by Δ_i^*.

When the moderating firm sets the passing level as ω, the fraction of the dedicated and the flexible agents are $F_\eta(\omega)$ and $1 - F_\eta(\omega)$, respectively. Similar to the *Benchmark* case, we can use Theorem 1 to determine the equilibrium revenue of agents. Our results in Theorem 1 suggests that the dedicated agents cannot generate any revenue if their service capacity exceeds the demand from class j, which occurs when $F_\eta(\omega) > \rho_j$. We show that the firm has to avoid these equilibrium outcomes resulting in zero revenue for the dedicated agents, and thus set $\omega \le F_\eta^{-1}(\rho_j)$. We also show that class i expects a higher reward from the flexible agents than class j does, and thus the firm prefers flexible agents to serve only class i customers when the service rate for class i is the higher one, i.e., when $\tau_i \ge \tau_j$. To ensure all flexible agents serve class i, the firm must choose a passing level that is greater than $F_\eta^{-1}(1-\rho_i)$ according to our results in Theorem 1. These two bounds on ω establish that the firm only needs to consider the interval of $[F_\eta^{-1}(1-\rho_i), F_\eta^{-1}(\rho_j)]$ while choosing the optimal passing level when $\tau_i \ge \tau_j$. We refer to this interval as the *dominating interval* and illustrate it below.

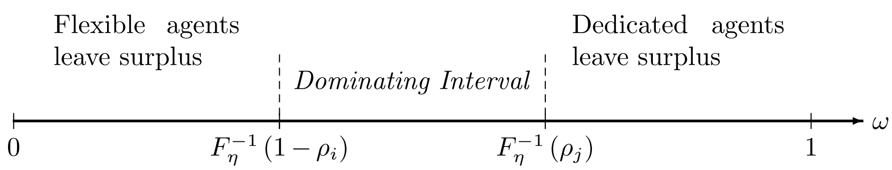


Figure 5 **The illustration of the dominating interval when the firm offers Exam $i \in \{A, B\}$ and $\tau_i \ge \tau_j$.**

In the dominating interval, we show that firm's revenue from each class increases by the service capacity allocated to this class. As the moderating firm cannot increase the service capacity for both classes simultaneously, it trades off between the gains from increasing the capacity for one class and the losses from decreasing the capacity for the other class. We show that the gains from increasing the service capacity for class j dominate the firm's losses from class i when the demand from class j is lower than a critical demand level $\overline{\rho}$. Thus, if $\rho_j < \overline{\rho}$, the moderating firm increases the service capacity for class j until the capacity meets the demand from this class. Then, it allocates the rest of the agents to class i. As a result of this skill-mix structure, all of the class j customers obtain service, whereas some customers from class i do not request service since there is not enough capacity to serve the entire class i. Similarly, the firm maximizes its revenue by serving all customers in class i when the demand from class i is lower than the critical level of $1-\overline{\rho}$. Finally, if the demands from both classes exceed the corresponding demand thresholds, the moderating firm maximizes its revenue at a passing level where the gain from increasing the capacity for one class is equal to the loss from decreasing the capacity for the other class. We formally present the above results in the following theorem.

THEOREM 3. *The optimal passing threshold when the firm offers Exam $i \in \{A, B\}$ and $\tau_i \geq \tau_j$ is*

$$\omega_i^* = \begin{cases} F_\eta^{-1}(\rho_j) & \text{if } \rho_j \leq \overline{\rho} \\ \overline{\omega}(\eta) & \text{if } 1-\rho_i < \overline{\rho} < \rho_j \\ F_\eta^{-1}(1-\rho_i) & \text{if } \overline{\rho} \leq 1-\rho_i, \end{cases}$$

where $\overline{\rho} = F_\eta(\overline{\omega})$ and $\overline{\omega}$ is the unique non-trivial solution for $\omega^{1/\eta} + \omega^\eta - 2\tau_i/\tau_j\omega = 0$.

The above theorem requires us to solve an analytically intractable fixed point problem to obtain an explicit form for the optimal passing level. Thus, it is not possible to use the above theorem to evaluate how much the firm benefits from an additional exam. To obtain insights for the revenue improvements gained in the *One-Test* case, we study the firm's problem under the limiting cases of the skill distribution, namely when the shape parameter η approaches infinity or one.

In our first limiting case, we let the shape parameter η grow to infinity. Proposition 2 shows that the skills of an agent become independently distributed when η approaches infinity. Furthermore, the marginal skill distribution, $F_\eta(\cdot)$, becomes a *Uniform* distribution in this limiting case. As the skill distribution approaches *Uniform*, we show that it is optimal for the firm to offer the exam related to the class with the highest service rate. In other words, the optimal exam is Exam A if $\tau_A \geq \tau_B$, and Exam B, otherwise. We also establish that the optimal exam threshold follows the structure proven in Theorem 3. Namely, supposing Exam A is the optimal exam for ease of explanation, the firm optimally sets aside enough capacity to serve all class B customers as long as their demand rate is below the critical level of $\tau_B/(2\tau_A)$. When the demand from class B is above $\tau_B/(2\tau_A)$, some of the customers from class B end up not getting the service. Furthermore, as long as the demand from class A, ρ_A, exceeds $1 - \tau_B/(2\tau_A)$, the firm distributes the service capacity in a proportional way that depends only on the service rates. Particularly, the firm allocates a higher fraction of agents to the class with the faster service times. To a certain extend, this allocation rule resembles the "*proportional allocation*" rule introduced in Cachon and Lariviere (1999). Once ρ_A is below $1 - \tau_B/(2\tau_A)$, the firm optimally allocates enough capacity to serve all customers from class A. By studying our first limiting case, we also show that offering an additional exam may improve the firm's revenues as much as 25% compared to the *Benchmark* case when the service rates are identical. More interestingly, the relative improvement in revenue surges as the service rates τ_A and τ_B diverge from each other. This result holds true because as the highest service rate increases, the firm's optimal revenues in the *One-Test* case rise whereas the revenues in the *Benchmark* case stays constant.

THEOREM 4. *If $\tau_i \geq \tau_j$ for $i \neq j \in \{A, B\}$, then we have that*

$$\lim_{\eta\to\infty} \omega_i^* = \begin{cases} \rho_j & \text{if } \rho_j \leq \frac{\tau_j}{2\tau_i} \\ \frac{\tau_j}{2\tau_i} & \text{if } 1-\rho_i < \frac{\tau_j}{2\tau_i} < \rho_j \\ 1-\rho_i & \text{if } \frac{\tau_j}{2\tau_i} \leq 1-\rho_i. \end{cases}$$

Furthermore, we have that

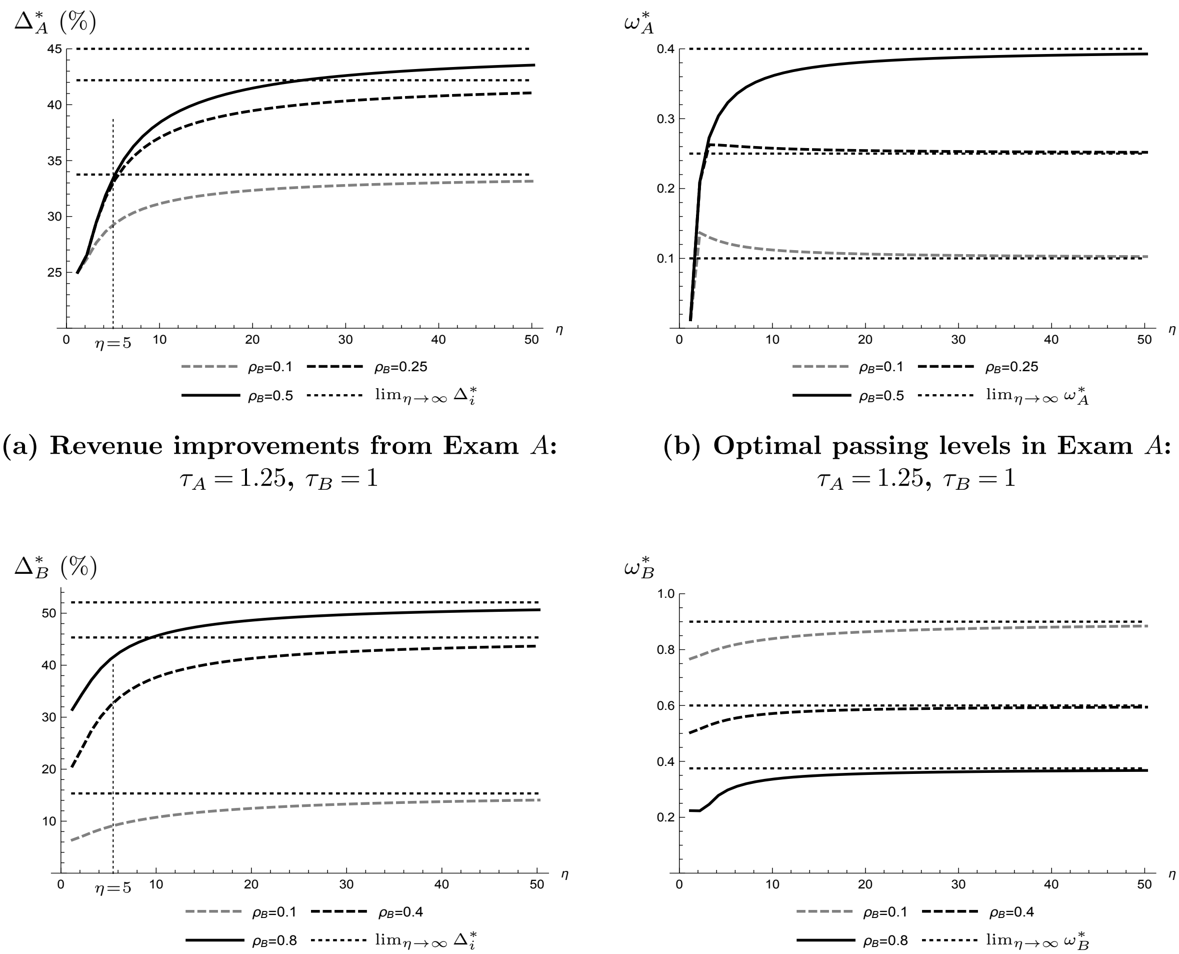
$$\lim_{\eta\to\infty} \Delta_j^* \leq \lim_{\eta\to\infty} \Delta_i^* = \begin{cases} (1-\rho_j)\left[(1+\rho_j)\frac{\tau_i}{\tau_j} - 1\right] & \text{if } \rho_j \leq \frac{\tau_j}{2\tau_i} \\ \frac{\tau_i}{\tau_j} + \frac{\tau_j}{4\tau_i} - 1 & \text{if } 1-\rho_i < \frac{\tau_j}{2\tau_i} < \rho_j \\ \rho_i\left[(2-\rho_i)\frac{\tau_i}{\tau_j} - 1\right] & \text{if } \frac{\tau_j}{2\tau_i} \leq 1-\rho_i. \end{cases}$$

Our second limiting case assumes the shape parameter η approaches one, and thus the agent skills become perfectly and positively correlated, as shown in Proposition 2. When the skills of an agent are perfectly correlated, we again show that the firm optimally offers the exam that screens the skill demanded by the customer class with the highest service rate. We also prove that under the optimal exam, the firm's revenues decline as the exam threshold increases in the *dominating interval* because the firm has an incentive to interchange the dedicated agents with the flexible ones. Hence, the firm sets the exam threshold at the level where the demand from class $i \in \{A, B\}$ is fully satisfied by the flexible agents when it is optimal to offer Exam i. We also establish that the firm can improve its revenues by offering one exam compared to the *Benchmark* case unless the service rates are identical. When the service rates are the same, i.e., when $\tau_A = \tau_B$, the dominating interval expands to include $\omega = 0$, which is the *Benchmark* case, and the firm's revenue is the same for any passing levels in the *dominating interval.* As the *Benchmark* case is a part of the *dominating interval* under perfectly correlated agent skills, the firm does not benefit from offering an additional test.

THEOREM 5. *If $\tau_i \geq \tau_j$ for $i \neq j \in \{A, B\}$, then we have that* $\lim_{\eta\to 1} \omega_i^* = F_1^{-1}(1-\rho_i)$. *Furthermore, we have that*

$$\lim_{\eta\to 1} \Delta_j^* \leq \lim_{\eta\to 1} \Delta_i^* = \frac{9}{4}\left[\frac{\tau_i}{\tau_j} - 1\right] \int_{F_1^{-1}(1-\rho_i)}^{1} 4s^2 \log(1/s) ds.$$

The above theorems establish the limiting behavior of the revenue improvements after offering an additional exam as the agent skills become independent or perfectly correlated. We perform a numerical study to illustrate the benefits from offering a skill test for the skill correlation level that are in-between these two extremes. In our numerical study, we consider different values for the demand rate of class B, as presented in Figure 6. We set the demand rate of class A to 1 to make sure that the total demand is always above 1 and $\rho_A \geq \rho_B$. Our numerical study shows that more than half of the asymptotical revenue improvement stated in Theorem 4 can be achieved even when η is as low as 5. It is worth noting that $\eta \geq 5$ implies a correlation coefficient of 0.5 or less. According to the histogram presented in Figure 4, this corresponds to almost 70% of exam pairs taken together at least 100 times in the data we collected from upwork.com. In our numerical study, we also show that benefits from offering one exam become more profound as the correlation between an agent's skills weakens, which happens as η increases.



(a) Revenue improvements from Exam A: $\tau_A = 1.25,\ \tau_B = 1$

(b) Optimal passing levels in Exam A: $\tau_A = 1.25,\ \tau_B = 1$

(c) Revenue improvements from Exam B: $\tau_A = 0.75,\ \tau_B = 1$

(d) Optimal passing levels in Exam B $\tau_A = 0.75,\ \tau_B = 1$

Figure 6 **The revenue improvements and optimal passing levels as η grows when the firm offers a-b) Exam A and c-d) Exam B.**

Considering the case of $\tau_A \geq \tau_B$, one of the major implications of Theorem 4 is that the optimal passing level is one of the end points of the *dominating interval* for large values of η when the lower demand rate is less than $\tau_B/(2\tau_A)$. The firm optimally chooses the end point which allocates enough capacity to match the demand from the class with the lowest demand rate. On the other hand, the optimal passing level is an interior solution when the lower demand rate is higher than $\tau_B/(2\tau_A)$. The firm divides the service capacity by allocating more agents for the class with the highest service rate. We illustrate these observations in Figure 6.b. Similar to the asymptotic behavior of the revenue improvement values, the optimal passing level also quickly converges to the limits established in Theorem 4. When $\tau_A \le \tau_B$, we have parallel results where only the critical level for the class B demand is altered from $\tau_B/(2\tau_A)$ to $1-\tau_A/(2\tau_B)$. Based on these observations, we propose a heuristic solution to the firm's problem under the *One Test* case: When the service rate for class A is the highest, the firm offers Exam A and sets the passing level to $F_\eta^{-1}(\rho_B)$ if $\rho_B < \tau_B/(2\tau_A)$ and $F_\eta^{-1}(\tau_B/(2\tau_A))$ otherwise. When the service rate for class B is higher, the firm offers Exam B and sets the passing level to $F_\eta^{-1}(1-\rho_B)$ if $\rho_B < 1-\tau_A/(2\tau_B)$ and $F_\eta^{-1}(\tau_A/(2\tau_B))$ otherwise. As a direct implication of Theorem 4, the heuristic solution we propose is asymptotically optimal, i.e. as $\eta \to \infty$, when the highest demand rate, ρ_A, is above $1-\tau_B/(2\tau_A)$, which is always true if $\tau_A \le \tau_B$. Furthermore, we numerically observe that our heuristic solution performs quite well. In fact, the revenue gap between the proposed heuristic and the optimal solution is below 1% even for low levels of the shape parameter η. We also want to note that when service rates are identical, our heuristic solution for large η and high demand rates is aligned with the skill test practices at upwork.com, where the passing levels are set to 1/2.

6.3. Two Skill Tests

As our final case, we now study the *Two-Tests* case. In this case, the firm sets strictly positive passing levels on both exams. Therefore, the firm's objective, in the *Two-Tests* case, is to maximize the revenue, $\Pi(\omega_A, \omega_B)$, given the constraints of $\omega_A > 0$ and $\omega_B > 0$. We denote the firm's optimal profit by Π^{**} and the optimal passing level in Exam $i \in \{A, B\}$ by ω_i^{**}. We also define the relative improvement in revenue from the *One-Test* case to the *Two-Tests* case as $\Pi^{**}/\max\{\Pi_A^*, \Pi_B^*\} - 1$ and denote it by Δ^{**}.

Unlike the previous case where the skill-mix structure is only one type, the marketplace in the *Two-Tests* case can be an *N-Network*, an *M-Network*, or a *V-Network*. This will make the firm's problem more challenging than the *One-Test* case. Therefore, similar to the *One-Test* case, we will focus on the limiting cases where the agent skills are (i) perfectly correlated and (ii) independent.

When the skills of an agent are perfectly and positively correlated, we find that the firm's revenues from the flexible agents do not depend on the passing level of the easiest exam (i.e., the exam with the loweest passing level). Hence, once the higheest passing level is fixed, the firm's only focus is its revenue from the dedicated agents. As the firm decreases the threshold of the easiest exam, there will be more dedicated agents but at the expense of lower average skills, which lead to lower equilibrium revenues. In other words, the firm has to tradeoff between the service capacity and the equilibrium revenue of agents. It turns out, the firm strictly prefers higher capacity, and thus always improves its revenue by lowering the passing levels of the easiest exam. In fact, the firm finds it optimal to eliminate the easiest exam completely, and thus prefers offering only one exam in a marketplace with perfectly and positively correlated agent skills even if a second exam is feasible.

In our second limiting case, we study the firm's problem when the agent skills are independent and *Uniformly* distributed. Similar to the perfect correlation case, the firm finds it profitable to decrease the passing level of the easiest exam to expand the service capacity. However, when the agent skills are independent, we find that the firm makes the easiest exam less challenging until the passing levels land on a critical curve. We show that the passing levels on this critical curve generate more revenue for the firm than the rest of the passing levels, thus we refer to it as the *dominating curve.* Figure 7 illustrates the *dominating curve.*

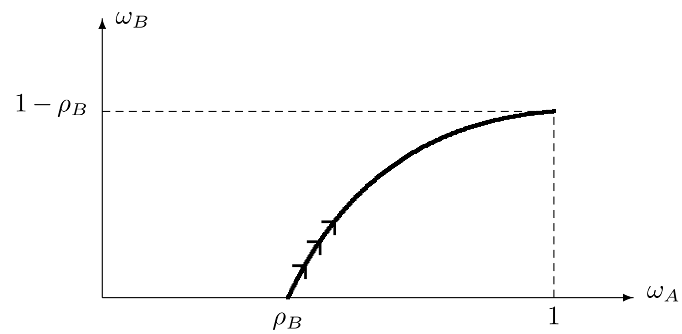


Figure 7 The dominating curve: Any (ω_A, ω_B) **on the dominating curve satisfies** $\omega_B = 1 - \rho_B/\omega_A$

As the firm moves on the *dominating curve* in the direction indicated by arrows (which could be achieved by increasing both passing levels), the firm faces a tradeoff: The average skill of a dedicated agent serving class B increases, which leads to higher revenues from class B, whereas the service capacity allocated to this class decreases, which means lower revenues from class A. We show that the gains from class B may outweigh the losses from class A as the firm move along the dominating curve. Thus, there may be an interior optimal solution on the *dominating curve*, and this implies that the firm is strictly better off running the second exam. However, we show that the relative improvement after the second exam cannot exceed 2.1% and is positive only when $\rho_B < \tau_B/(2\tau_A)$. In other words, offering the second exam does not bring any extra benefit to the firm when $\rho_B \geq \tau_B/(2\tau_A)$. We formally present these findings in the following theorem.

THEOREM 6. 1. *If* $\rho_B \geq \tau_B/(2\tau_A)$, *then we have that* $\lim_{\eta\to\infty} \Delta^{**} = 0$.

2. *If* $\rho_B < \tau_B/(2\tau_A)$, *then we have that* $\lim_{\eta\to\infty} \Delta^{**} \leq 2.1\%$.

3. $\lim_{\eta\to 1} \Delta^{**} = 0$*, i.e., offering an additional test does not improve the revenue of the firm when* η *approaches 1.*

The above theorem provides an upper bound for the benefits obtained from the second exam as $\eta \to \infty$. The exact expression, which is omitted for brevity, can be found in the proof of the theorem. We illustrate the relative improvement after the second exam as the skills become independent, $\lim_{\eta\to\infty} \Delta^{**}$, in Figure 8. As this figure shows, the improvements exceed 1% only for a small interval of the demand rate of class B. It is also important to note that the benefits from the second exam diminishes as the discrepancy between the service rates increases. This is in contrast with our findings regarding the benefits from the first exam.

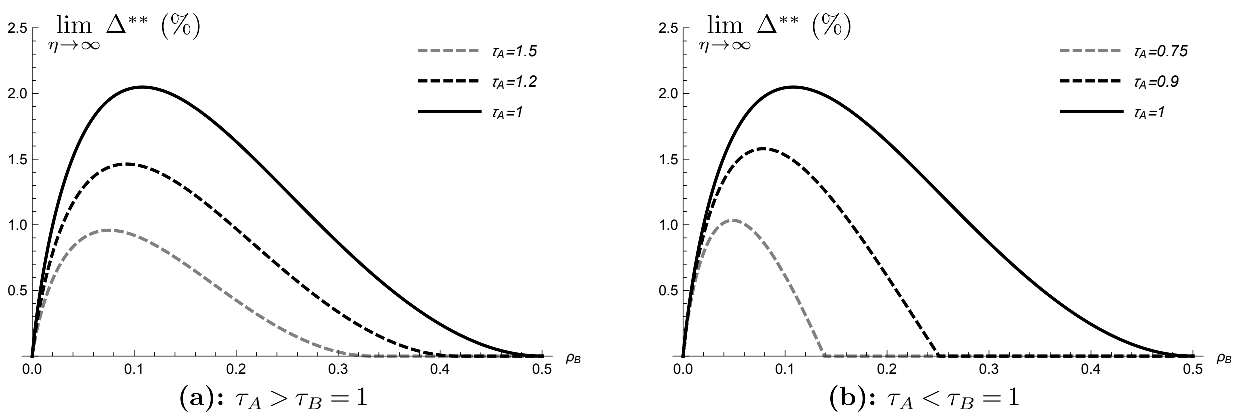


Figure 8 Relative revenue improvements from the One-Test case to the Two-Tests case when $\rho_A = 1$, $\eta = \infty$.

Similar to the *One-Test* case, we perform a numerical study to illustrate the limiting behavior proven in Theorem 6 for the case of identical service rates. We consider identical service rates in our numerical study since it is the case where the second exam yields the highest benefits. Unlike the Δ^*, Δ^{**} is not always increasing in the shape parameter η as shown in Figure 9. This non-monotone structure occurs because Δ^{**} captures the revenue improvements compared to the *One-Test* case. We verified that the relative improvement in revenue from the *Benchmark* case to the *Two-Tests* case increases as the correlation between agent skills declines (i.e., as η increases).

6.4. The Optimal Number of Tests

In the previous subsections, we study the firm's problem by fixing the number of skill tests it offers. We show that the revenue improvements from offering skill tests highly depend on the correlation

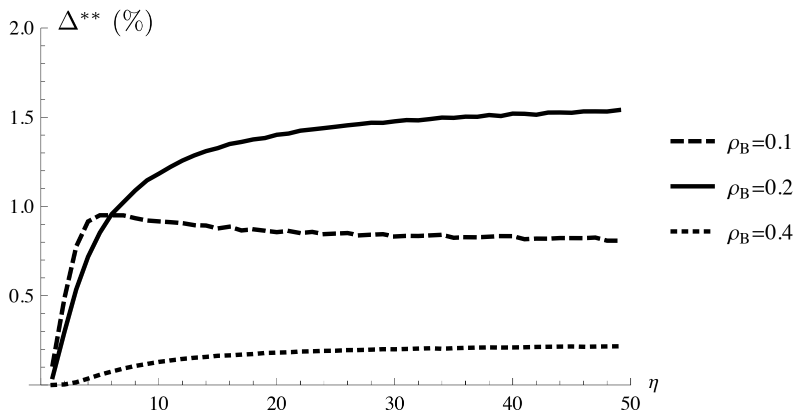


Figure 9 Relative revenue improvements from the One-Test case to the Two-Tests case as a function of η when $\tau_A = \tau_B$.

between the skills of an agent. Specifically, we find that offering a skill test slightly improves the firm's revenue when the agent skills are highly correlated. On the other hand, if the correlation between skills is negligible, the first additional test can lead to substantial revenue improvements. However, the second exam cannot generate similar levels of significant revenue improvements. We also numerically show that the firm benefits from the tests more as the skills become less correlated.

When the firm does not incur any costs for preparing and offering a skill test, the *Two-Tests* case is naturally the best option for the firm because the firm can set the passing levels to zero if needed. In the absence of skill screening costs, our findings help the moderating firm to choose the optimal level of difficulty in each exam. For instance, our results suggest that the firm should let all candidate agents pass both exams when skills are perfectly correlated, while setting strictly positive passing levels in both exams is optimal if the skill correlation can be disregarded. However, it is not unrealistic to consider a preparation and implementation cost associated with the skill tests. Then, the firm can use our findings to choose the optimal number of tests to be offered. As we show in the following corollary, letting C be the cost for offering a skill test, we can find a critical level of shape parameter $\underline{\eta}$ and show that the firm optimally offers no tests when $\eta < \underline{\eta}$. Similarly, we can show that it is optimal for the firm to offer only one test for sufficiently low levels of skill correlation as long as C is greater than a small percentage (less than 1.7%) of the total revenue.

Corollary 1. *For any $C > 0$, there exists a $\underline{\eta}$ such that it is optimal for the firm to offer zero tests for any $\eta < \underline{\eta}$. Furthermore, if $C/\Pi^{**} > \lim_{\eta\to\infty} \Delta^{**}$ and $C/\Pi^{o} < \lim_{\eta\to\infty} \Delta^{*}$, then there exists a $\overline{\eta}$ such that it is optimal for the firm to offer only one test for any $\eta > \overline{\eta}$.*

In the above corollary, we consider a fixed cost for offering skill test. One can also envision a marketplace where the cost associated with offering tests is proportional to the revenue generated in the marketplace. For example, skill tests may discourage service providers from participating, so that the service provision capacity may be affected by the skill tests. To be specific, we can consider a marketplace where a predetermined portion of agents choose to leave the marketplace for each test the firm offers. Then, similar to the above corollary, we can find a critical level of η under which the firm optimally offers no tests. We can also show that offering only one test is optimal as long as the skill correlation is sufficiently low.

7. Conclusion

In this paper, we study a marketplace in which many small service providers compete with each other in providing service to two groups of self-interested customers. Service providers are distinguished with respect to their service skills, and each group of customers has different needs. An important aspect of these marketplaces that our model captures is that customers cannot learn the skills of a provider before the completion of the service. However, the moderating firm, which sets up the marketplace, may help customers by providing them with further information about the ability of candidate agents through a skill screening mechanism. Such a screening mechanism consists of skill tests determining whether or not a candidate agent is eligible to serve customers. Skill screening also helps the firm to create different skill-mix structures in the marketplace.

The main focus of this paper is to gain insights about how the moderating firm can use skill screening as a tool to maximize its revenue. Hence, we study a problem where the firm can offer two skill tests and choose passing levels in the tests it offers. As the online marketplaces we review usually attract service providers with complementary skills, we consider a family of agent skill distributions where the correlation between skills ranges from 1 (perfect correlation) to 0 (independence). We show that the level of correlation between agent skills plays a crucial role on how the firm uses the skill tests. For instance, when the agent skills are highly correlated and customers are homogenous in their average service time, skill screening can hardly improve the firm's revenue. On the other hand, we show that the firm starts to obtain considerable benefits from skill screening as the skill correlation softens. We also show that the revenue improvements due to skill screening surge as the customers require different processing times. It turns out that the firm does not need to run both of the exams to achieve these high levels of benefits from testing.

As we mentioned before, one can view skill screening as a tool to regulate the marketplace. Thus, our results also shed light on the relationship between the level of skill correlation and how much intervention the marketplace requires. In particular, we show that the higher the correlation between the skills of an agent is, the less regulation/intervention the firm needs. Furthermore, we recommend that moderating firms not use both of the available exams when regulation is necessary.

In this paper, we study a marketplace where the service capacity is scarce and customers' outside utility is normalized to zero. Our key findings, which explain the relationship between skill screening and correlation, would continue to hold when the capacity is ample and customers have a positive outside utility. In the case of ample capacity (i.e., when $\rho_A + \rho_B < 1$), the firm has to choose strictly positive passing levels because otherwise, service providers would charge very low prices due to intensified competition. Similarly, when the customers have a strictly positive outside utility, it is profitable for the firm to increase the passing levels to match this outside option. In fact, relaxing these two assumptions will shift the origin of the feasible passing level space outwards from $(0,0)$ to $(\underline{\omega}_A, \underline{\omega}_B)$, where both passing levels are positive. Thus, the firm always offers two tests, even when they are costly. However, the remaining important question is how much benefit the firm can obtain when it further intervenes in the marketplace by making the skill tests more comprehensive than the shifted origin. After relaxing the aforementioned assumptions, we can show that the firm obtains significant benefits from making one of the exams more comprehensive compared to the new origin $(\underline{\omega}_A, \underline{\omega}_B)$. Similar to our results in Section 6, this implies that intervening in the marketplace via only one exam can lead to sizable revenue gains for the firm. However, setting the passing levels higher than $(\underline{\omega}_A, \underline{\omega}_B)$ in both exams only leads to small gains, which again establishes that running the second exam is not advisable for the moderating firm when skill screening is costly.

The jobs that we study in this paper require only one skill, and agents need to reveal their test results in order to attract a reasonable amount of customer demand. Considering the large-scale nature of the online marketplaces, some marketplaces may face customers requesting a combination of skills. It is also possible that agents who may choose not to reveal their test results yet attract sufficient demand. Future work can explore these more complex marketplace models and derive new insights about how the marketplaces can utilize skill tests.

Acknowledgments

The authors are grateful to Martin Lariviere, Michael Pangburn, and Laurens Debo for their helpful suggestions. The authors also thank the review team for their valuable comments, which substantially improved this paper.

References

Allon, G., A. Bassamboo, E. B. Çil. 2012. Large-scale service marketplaces: The role of the moderating firm. *Management Sci.* **58**(10) 1854–1872.

Allon, G., A. Federgruen. 2007. Competition in service industries. *Operations Research* **55**(1) 37–55.

Allon, G., I. Gurvich. 2010. Pricing and dimensioning competing large-scale service providers. *Manufacturing & Service Operations Management* **12**(3) 449–469.

Amihud, Y., H. Mendelson. 1980. Dealership market: Market-making with inventory. *Journal of Financial Economics* **8**(1) 31–53.

Armony, M., C. Maglaras. 2004. On customer contact centers with a call-back option: Customer decisions, routing rules and system design. *Operations Research* **52**(2) 271–292.

Bassamboo, A., R. S. Randhawa, J. A. Van Mieghem. 2010. Optimal flexibility configurations in newsvendor networks: Going beyond chaining and pairing. *Management Sci.* **56**(8) 1285–1303.

Bernheim, B. D., B. Peleg, M. Whinston. 1987. Coalition-proof nash equilibria i: Concepts. *Journal of Economic Theory* **42**(1) 1–12.

Biais, B., L. Glosten, C. Spattt. 2005. Market microstructure: A survey of microfoundations, empirical results, and policy implications. *Journal of Financial Markets* **8**(2) 217–264.

Bish, E. K., Q. Wang. 2004. Optimal investment strategies for flexible resources, considering pricing and correlated demands. *Operations Research* **52**(6) 954–964.

Burdett, K., D. Mortensen. 1998. Wage differentials, employer size, and unemployment. *International Economic Review* **39**(2) 257–273.

Cachon, G., P. T. Harker. 2002. Competition and outsourcing with scale economies. *Management Sci.* **48**(10) 1314–1333.

Cachon, G. P., M. A. Lariviere. 1999. An equilibrium analysis of linear, proportional and uniform allocation of scarce capacity. *IIE Transactions* **31**(9) 835–849.

Callen, J. L., B. Sarath. 1995. Risk aversion, generalized correlation and investment in manufacturing capacities. *Journal of Operations Management* **12**(3-4) 225–238.

Chen, H., Y. Wan. 2003. Price competition of make-to-order firms. *IIE Transactions* **35**(9) 817–832.

Chen, Y., C. Maglaras, G. Vulcano. 2008. Design of an aggregated marketplace under congestion effects: Asymptotic analysis and equilibrium characterizatio. Working Paper.

Chod, J., N. Rudi. 2005. Resource flexibility with responsive pricing. *Operations Research* **53**(3) 532–548.

Fine, C. H., R. M. Freund. 1990. Optimal investment in product-flexible manufacturing capacity. *Management Science* **36**(4) 449–466.

Garman, M. 1976. Market microstructure. *Journal of Financial Economics* **3**(3) 257–275.

Goyal, M., S. Netessine. 2007. Strategic technology choice and capacity investment under demand uncertainty. *Management Science* **53**(2) 192–207.

Gupta, D., Y. Gerchak, J. A. Buzacott. 1992. The optimal mix of flexible and dedicated manufacturing capacities: Hedging against demand uncertainty. *International Journal of Production Economics* **28**(3) 309–320.

Halfin, S., W. Whitt. 1981. Heavy-traffic limits for queues with many exponential servers. *Operations Research* **29**(3) 567–588.

Ho, T., H. Stoll. 1983. The dynamics of dealer markets under competition. *Journal of Finance* **38**(4) 1053–1074.

Maglaras, C., A. Zeevi. 2003. Pricing and capacity sizing for systems with shared resources: Scaling relations and approximate solutions. *Management Science* **49**(8) 1018–1038.

Manning, A. 2004. Monopsony and the efficiency of labour market interventions. *Labour Economics* **11**(2) 145–163.

Michaelides, M. 2010. Labour market oligopsonistic competition: The effect of worker immobility on wages. *Labour Economics* **17**(1).

Moreno, D., J. Wooders. 1996. Coalition-proof equilibrium. *Games and Economic Behavior* **17**(1) 80–112.

Naor, P. 1969. The regulation of queue size by levying tolls. *Econometrica* **37**(1) 15–24.

Ray, I. 1996. Coalition-proof correlated equilibirum: A definition. *Games and Economic Behavior* **17**(1) 56–79.

Upwork Help Center. 2016. Help center: Skills tests. URL `https://support.upwork.com/hc/en-us/articles/211063198-Skills-Tests`.

Van Mieghem, J. A. 1998. Investment strategies for flexible resources. *Management Science* **44**(8) 1071–1078.

Whitt, W. 2006. Fluid models for multiserver queues with abandonments. *Operations Research* **54**(1) 37–54.

Zeltyn, S., A. Mandelbaum. 2005. Call centers with impatient customers: Many-server asymptotics of the M/M/n+G queues. *Queueing Systems* **51**(3) 361–402.

APPENDICES

Appendix A: Table of Notation

$i \in \{A, B\}$	Indicator for class or Exam i.	Λ_i	Rate of the arrival process for class i.
τ_i	Rate of the service process for class i.	c_i	Waiting cost per unit time for class i.
$\underline{u}$	The utility from the outside option.	S_i	Value that an agent's service generates for a class i.
$f_{A,B}(\cdot,\cdot)$	Joint probability density function for S_A and S_B.	ρ_i	Demand-supply ratio of class i, $\rho_i \equiv \Lambda_i/(\tau_i k)$.
$\alpha_i(\omega_A, \omega_B)$	Fraction of dedicated agents for class i given the passing levels (ω_A, ω_B).		
$\alpha_F(\omega_A, \omega_B)$	Fraction of flexible agents given the passing levels (ω_A, ω_B).		
$R_i(\omega_A, \omega_B)$	Expected reward for a class i customers from dedicated agents given the pair of passing levels (ω_A, ω_B).		
$R_{iF}(\omega_A, \omega_B)$	Expected reward for a class i customers from flexible agents given the pair of passing levels (ω_A, ω_B).		
$(r_{i_n}, y_{i_n}, t_{i_n})$	Pricing and service strategies of the agents in sub-pool i_n. r_{i_n} is the net-reward. t_{i_n} is portion of the service capacity that agents allocate to class i. $y_{i_n}k$ is the number of agents in the sub-pool.		
N_i	Number of different sub-pools serving class i	$\overline{\alpha}_i$	Fraction of total service capacity available for class i, $\overline{\alpha}_i \equiv \sum_{n=1}^{N_i} t_{i_n} y_{i_n}$.
D_i	Fraction of class i customers requesting service.	$W(\rho)$	Expected waiting time in the generic $M/M/s$ with the arrival rate of $s\rho$ and the service rate of 1.
$P_\ell(\mathbf{r}, \mathbf{y}, \rho)$	Probability with which a customer is served by an agent in sub-pool ℓ in the generic $M/M/s$ with the arrival rate of $s\rho$ and the service rate of 1		
$\sigma_\ell(\mathbf{r}, \mathbf{y}, \rho)$	Utilization of agents in sub-pool ℓ in the generic $M/M/s$ with the arrival rate of $s\rho$ and the service rate of 1.		
η	Shape parameter for the family of joint skill distributions we consider in Section 6.	$F_\eta(\cdot)$	Marginal distribution of each agent skill in Section 6.
$\mathbf{E}[S_\eta]$	Averages value that an agent's service generates for each classes without testing in Section 6.	Π^o	Revenue of the marketplace in the *Benchmark* case.
Π_i^*	Optimal revenue of the marketplace in the *One-Test* case under Exam i.	Δ_i^*	Relative improvement in revenue from the *Benchmark* case to the *One-Test* case.
Π^{**}	Optimal revenue of the marketplace in the *Two-Test* case.	Δ_i^{**}	Relative improvement in revenue from the *One-Test* case to the *Two-Test* case.

Table 2 **Descriptions of the frequently used notation.**

Appendix S: Online Supporting Document

Additional supporting information may be found in the online supplement of this article:

Appendix S.1: Proofs in Section 5

Appendix S.2: Proofs in Section 6