Asymmetric information and economies-of-scale in service contracting

Mustafa Akan, Bariş Ata,
Martin A. Lariviere

September 5, 2007
COSM-07-001

Working Paper Series

Center for Operations and Supply Chain Management

Kellogg School of Management

Asymmetric information and economies-of-scale in service contracting

Mustafa Akan • Barış Ata • Martin A. Lariviere
Kellogg School of Management, Northwestern University

September 5, 2007

Abstract

We consider a contracting problem in which a firm outsources its call center operations to a service provider. The outsourcing firm (which we term the originator) has private information regarding the rate of incoming calls. The per-call revenue (or margin) earned by the firm and the service level depend on the staffing decisions by the service provider. Initially, we restrict attention to pay-per-call contracts under which the parties contract on a service level and a per-call fee. The service provider is modeled as a multi-server queue with a Poisson arrival process, exponentially distributed service times and customer abandonment. We assume that the service provider's queue is large enough such that the economically sensible mode of operation for staffing it is the Quality-and-Efficiency-Driven regime, which allows tractable approximations of various performance metrics. We first consider a screening scenario with the service provider offering a contract to the originator. Due to the statistical economies of scale phenomenon observed in queueing systems, the allocation of the originator with higher arrival rate is distorted, which reverses the typical "efficiency at top" result present in the literature on monopolist screening. We then consider the alternative scenario with the originator offering a contract to signal her information and show that the service level of the high volume firm is again distorted. The introduction of a fixed payment ameliorates distortions from first-best and may eliminate them.

1 Introduction

The last twenty years have seen an increasing use outsourcing. Beginning in manufacturing, firms have made greater and greater use of others to do work for them. Over time, this has moved from purchasing simple commodity parts to having suppliers provide complex parts and subassemblies. For its new 787 Dreamliner, Boeing is counting on outside firms to deliver doors, landing gear, and even entire wings (Niezen & Weller (2006)). In the auto industry, it is not uncommon for purchased materials to account for more than half the cost of making a car.

Outsourcing, however, is not just for manufacturing anymore. Recent years have seen a growing trend in outsourcing services. Such outsourcing began with simple ancillary activities such as janitorial services but has grown to include the complete outsourcing of entire business processes (such as order taking) and departments (such as information technology). This has become a big business. TPI, a consulting firm that specializes in sourcing, estimates that new business process outsourcing (BPO) contracts for the first half of 2007 exceeded $30 billion in value, which is actually a decrease from earlier years, cf. Munoz (2007). TPI's estimates do not include government contracts, deals under $50 million, or contracts renewed without the help of outside consultants. Taking a broader view, Cohen & Young (2006) estimate that the BPO market will exceed three-quarters of a trillion dollars by 2008.

Given the growth and importance of service outsourcing, it is worth studying the similarities and differences between outsourcing the delivery of services and the production of physical goods. The reliance on suppliers to deliver components and subassemblies has been studied extensively in the supply chain management literature. In particular, researchers have studied a wide variety of contracts that govern the supplier-buyer relationship and how these vary with model parameters, available information, and relevant decisions. (See Cachon (2003), for detailed review.) This work generally takes a simple operational view of the supply chain in order to illuminate the role of economic incentives. Here, we take a similar tack, looking at a basic model in order to show how the move from a supply chain to a service settings alters the nature of optimal contract.

Specifically, we consider a firm (which we term the originator) outsourcing its inbound call center to a service provider. Admittedly, call center outsourcing is only one slice of the overall BPO market. However, it is a significant slice. Beasty (2005) estimates that the US call center outsourcing market will reach £23.9 billion by 2008 while Beasty (2006) reports that there are over 150 outsourcing vendors for call centers in North America. Further, looking at call centers allows us to focus on a business with well-documented and understood definitions of service quality. In other parts of the BPO markets, how to define and measure quality and performance is often a stumbling block to developing successful commercial partnerships (Taylor & Tofts (2006)).

Call center contracts also take a variety of form. Hourly charges based on the number of call center agents available to take calls for a client are common in the market.[1] Another alternative is a fee for each minute that an agent is engaged with a caller. Yet another possibility is a charge per call

[1] Much of the information on typical call center contracts comes from Centris Information Systems (2007) and conversations with a Centris executive.

handled. (If the distribution of call durations is well understood, per-call and per-minute-engaged charges are equivalent in the long run.) Because the typical call center pays its workers an hourly wage, hourly charges and payments tied to call volume have different risk implications. Hourly charges insure the call center covers its costs even if agents are idle. When payments are tied to the number of calls coming in, the service provider may not earn enough to recoup its staffing costs. As a consequence, per-call payments may be combined with some minimum volume requirements. Centris Information Systems of Longview, Texas, has contracts pairing a per-minute charge with a minimum utilization level. Accolade Support, an Albuquerque-based vendor, offers packages in which buyers commit to purchasing a minimum number of minutes with additional minutes of service being provided at higher prices.[2] Beyond ongoing charges, call originators typically pay a fixed amount at the start of the contract to cover agent training.

The amount charged for call center services, of course, depends on a number of factors. Complex services (such as order taking or technical support) require more highly skilled agents than simpler tasks (such as lead generation) and thus cost more. The agents serving an originator's may be dedicated to that originator or cross-trained to handle calls from multiple sources. Complex calls favor dedicated agents to reduce training requirements. The targeted service level whether measured by waiting time targets or abandonment rates also matters, with higher service levels increasing costs. This last point is crucial to our analysis. Higher service levels are more costly but just how much more costly depends on the call volume. Call centers naturally exhibit of economies of scale, so it is cheaper on a per call basis to provide good service when the call volume is high. This is reflected in pricing. Accolade Support has a base charge of 83.2¢ per minute when a customer commits to buying 600 minutes per month. That drops to 76.9¢ per minute when the customer commits to 6,500 minutes per month and to 70.4¢ per minute at 27,000 minutes. A call originator thus has an incentive to exaggerate his call volume in order to secure a lower cost, a better service level, or some combination of both.

We explore these issues in a simple contracting setting. The originator needs to hire a provider to service a flow of calls. Calls are revenue generating and better service results in both fewer callers abandoning as well as higher revenues per call. The originator can be one of two types. A high type, as the name implies, anticipates a higher average call volume than a low type. The originator knows her type with certainty but the service provider only has a prior over the two possibilities. We consider two sequences of play. In the first, the service provider moves first and offers a menu

[2] See www.accoladesupport.com/NewFiles/Accolade%20Support-Pricing.pdf.

of contracts. This raises the possibility that the provider may *screen* the types of customers. In the second, the originator moves first, proposing a contract that might *signal* his private.

Related problems of asymmetric demand information have been studied in the literatures on franchising (Desai & Sirinivasan (1995)), channel management (Chu (1992)), and supply chain management (Cachon & Lariviere (2001)). Like much of this work, we focus on terms of trade that are variations of two part tariffs. We first suppose that contracts consist of a per call fee and a promised service level (measured by the abandonment rate) and later add a lump sum payment as well. While our contracts are standard (both in the industry and the academic literature) our model of the market and the production process deviate from past academic work on contracting with asymmetric demand information. Most work in this area suppose linear production costs and either deterministic demand (given full information) or a newsvendor formulation. Our analysis is built around a queuing model to capture the natural economies of scale in running a call center and to endogenize service measures such as abandonments and average delays.

The switch to a queuing model has a nontrivial impact on the nature of contracting. A basic problem in mechanism design considers a seller offering multiple levels of quality to customer segments that differ in their willingness to pay for quality. Providing higher quality is more costly. This leads to the standard results that the most favorable type (i.e., customer segment which is more profitable) receive an efficient level of quality and captures an information rent. The least favorable type (i.e., customer segment which is less profitable) receive an inefficient level of quality and are driven to indifference. (See Salanié (1997).)

In our setting, these results are reversed[3]. Suppose that the service provider offers the contract. Here, the high volume provider is the favorable type. If the provider could verify the originator's type, he would provide a high volume originator with a higher quality service (i.e., a lower abandonment rate) in part because the marginal cost of providing good service is decreasing in the originator's call volume. When the provider cannot verify the originator's volume, he offers a menu that distorts the quality provided to the high type. The high type is offered an inefficiently low abandonment rate and is indifferent to accepting the contract. The low type, in contrast, is offered the efficient abandonment rate and garners an information rent. These results are slightly modified if the provider also imposes a lump sum payment; if the difference in arrival rates is sufficiently

[3]We are not the first to highlight how the economies of scale inherent in queueing systems alter the economic incentives. Cachon & Harker (2002) examines the impact of queueing economies of scale in a duopoly market. We, however, focus on a vertical as opposed to a horizontal relationship.

large, the high type may receive the efficient level of service but her abandonment rate is still distorted otherwise.

When the order of play is reversed and the originator offers the contract, a high volume originator signals her information by demanding a lower abandonment rate and adjusting her price. Thus, we conclude that asymmetric information results in inefficiently low abandonment rates for high call volumes regardless of who offers the contract. Again, this is modified when a fixed fee is included, which allows a high type to signal her information without distorting her desired service level.

Contract theory provides the framework for analyzing the strategic interactions among agents that arise from informational asymmetries, see Salanié (1997) for an introduction to agency models. Hassin & Haviv (2002) provides a survey for the strategic issues arising in queueing systems. The literature on service contracting for call center operations is thin. Aksin, de Vericourt & Karaesmen (2006) considers a service provider who is faced with an uncertain volume and can outsource all or part of the calls. The authors look at the impact of contract terms on the capacity planning and the nature of the work to be outsourced. In a newsvendor framework, they examine whether the firm should outsource its base load of calls or its peak calls. Ren & Zhou (2006) studies contracting in a service supply chain and analyzes contracts that can induce the call center to choose optimal staffing and effort choices. Ren & Zhang (2007) examines service outsourcing contracts when the service provider's cost structure is private information. Hasija, Pinker & Shumsky (2006) considers a variety of contracts for call center outsourcing and shows how different combinations of the contract features make the firm to better manage vendors when there is information asymmetry about worker productivity. Milner & Olsen (2007) considers a call center with both contract and non-contract customers, which gives priority to the contract customers only in the off-peak hours. The authors show that under contracts on the percentile of delay, this is the rational behavior on behalf of the call center. They also propose other novel contracts that eliminate such undesirable behavior from the perspective of the contract customers.

The rest of the paper is structured as follows: Section 2 introduces the model and analyzes the full-information case. Sections 3 & 4 examine, respectively, screening and signalling with pay-per-call contracts. Section 5 adds a fixed fee to the contract and section 6 concludes. Proofs are relegated to Appendices A through C throughout the paper.

2 The Model

Consider a contracting problem in which an originator outsources her call center operations to a service provider (also referred to as the call center). We assume that the calls are revenue generating and lead to the originator capturing a margin of m. Hence, our model is appropriate for an originator outsourcing order taking as opposed to one outsourcing technical support.

For simplicity, the service system is modeled as a multi-server queue with a Poisson arrival process and exponentially distributed service times. The arrival rate is λ, which for the moment is assumed commonly known. Without loss of generality, the mean service time is one. Each customer waiting to be served may abandon, and time-to-abandon is exponentially distributed with rate η. In other words, using the terminology that is standard in queueing theory, we model the call center as a $M/M/N+M$ queue.

We assume that λ is large, i.e., $\lambda \gg 1$. Thus, the approximation results of Garnett, Mandelbaum & Reiman (2002) for large call centers accurately captures the queueing dynamics in our setting. Besides developing approximations of various performance metrics for large call centers, Garnett et al. (2002) also argues that economic considerations require managing such large call centers in a *Quality-and-Efficiency-Driven* regime. That, in turn, leads to staffing decisions based on a square-root rule. Specifically, the number of call-center agents N prescribed by the square-root staffing rule is

$$N = \lambda + \beta\sqrt{\lambda}. \tag{1}$$

Given Poisson arrivals, $\sqrt{\lambda}$ is the of standard deviation of demand arriving per unit of time. β is then the system's excess capacity measured in units of the standard deviation of demand per unit of time. We will refer to β as the standardized excess capacity.

Garnett et al. (2002) shows that the standardized excess capacity β captures the impact of capacity decisions on various performance metrics. Consequently, we assume that per-call revenue (or margin) captured by the originator $m(\cdot)$ is a function of the system's standardized excess capacity. We assume $m(\beta)$ is increasing and concave in β. This is in line with the literature that assumes customers respond to the "full" price of the service and are concerned with explicit monetary charges as well as implicit non-monetary delay costs (Hassin & Haviv (2002)). A higher excess capacity results in better service, reducing a customer's non-monetary expense and allowing for a higher explicit monetary price.

The staffing cost for the call center is k per agent per unit of time. The originator and provider do not, however, contract explicitly on the number of agents. Rather, the parties contract on some service level and a per-call fee, which are commonly used terms of trade, cf. Hasija et al. (2006). Specifically, we assume the contract specifies an abandonment probability. Thus, we restrict attention to contracts of the form (α, c), where α is the agreed upon abandonment probability and c is the fee to be paid by the originator to the service provider per answered call.

We borrow the following approximation from Garnett et al. (2002) as a tractable model of call-center operations. The abandonment probability is given[4] by

$$\mathbb{P}(\text{Abandonment}) = \frac{\Delta(\beta)}{\sqrt{\lambda}}, \tag{2}$$

which provides a good approximation for moderate to large values of λ. Moreover, we assume that $\Delta(\cdot)$ is a convex decreasing function of the standardized excess capacity β. Garnett et al. (2002) provides an explicit formula for $\Delta(\cdot)$ in terms of the hazard rate function of a standard normal random variable, and indeed our assumptions on $\Delta(\cdot)$ are satisfied for moderate to large values of β, which is precisely the regime we are interested in.

For our purposes, the most important feature of the approximation (2) is that it captures the statistical economies of scale phenomenon one expects in large call centers. To elaborate on this, define the standardized excess capacity $\beta(\alpha, \lambda)$ needed to provide the abandonment rate α for a given arrival rate λ as follows.

$$\beta(\alpha, \lambda) = \Delta^{-1}(\alpha\sqrt{\lambda}), \tag{3}$$

where $\Delta^{-1}(\cdot)$ is the inverse of $\Delta(\cdot)$. Then for a given α, the corresponding standardized excess capacity $\beta(\alpha, \lambda)$ is decreasing in the arrival rate. In other words, a larger call center is more efficient and can provide better service for a given level of standardized excess capacity, which is precisely due to the economies of scale phenomenon.

For technical simplicity, we assume that the expected margin per incoming call is increasing and concave in the standardized excess capacity. This is formalized in the following assumption.

Assumption 1 $m(\beta)\left(1 - \frac{\Delta(\beta)}{\sqrt{\lambda}}\right)$ *is concave increasing in* β.

Finally, we assume the following holds.

[4] To be more specific, Garnett et al. (2002) proposes the approximation $\mathbb{P}(\text{Abandonment}) = \Delta(\beta)/\sqrt{N}$ which is equivalent to the approximation (2) asymptotically; and both are justified through the same limiting argument.

Assumption 2 *For* $\alpha_0 > \alpha_1$ *and* $\kappa > 1$,

$$\begin{aligned} & (1-\alpha_1)\, m\left(\Delta^{-1}\left(\alpha_1\sqrt{\kappa\lambda}\right)\right) - (1-\alpha_0)\, m\left(\Delta^{-1}\left(\alpha_0\sqrt{\kappa\lambda}\right)\right) \\ > \; & (1-\alpha_1)\, m\left(\Delta^{-1}\left(\alpha_1\sqrt{\lambda}\right)\right) - (1-\alpha_0)\, m\left(\Delta^{-1}\left(\alpha_0\sqrt{\lambda}\right)\right). \end{aligned}$$

As the service level increase (i.e., as α falls), profit per incoming call increases for two reasons. First, a caller is less likely to abandon. Second, a higher service level requires greater excess capacity, increasing margin per answered call. This assumption impose additional structure on how the profit per call increases. We are assuming that an increase in the service level is worth more on a per call basis as the arrival rate increase. This is reasonable because the higher volume system starts from a lower level of excess capacity (i.e., $\beta(\alpha_0, \kappa\lambda) < \beta(\alpha_0, \lambda)$ for $\kappa > 1$). It is also similar to the single crossing property generally imposed in the literature (Salanié (1997)).

2.1 The optimal full-information service level

To establish a benchmark, we now examine the optimal contract under symmetric information. Since the per-call payment by the originator to the call center is just a monetary transfer, the problem reduces to choosing the efficient abandonment rate given λ to maximize the system-wide profits. Hence, the efficient (also called the first-best) levels of abandonment probabilities are obtained by solving the following problem. Choose the abandonment probabilities α so as to

$$\text{maximize} \;\; m(\beta(\alpha, \lambda))\lambda(1-\alpha) - k(\lambda + \sqrt{\lambda}\beta(\alpha, \lambda)). \tag{4}$$

Recall that (3) provides a one-to-one correspondence between the abandonment probability and the associated standardized excess capacity for each type. Thus, (4) can equivalently be viewed as a problem of choosing the efficient level of standardized excess capacity for each type. That is, choose the standardized excess capacity levels β so as to

$$\text{maximize} \;\; m(\beta_H)\lambda_H\left(1 - \frac{\Delta(\beta_H)}{\sqrt{\lambda_H}}\right) - k(\lambda_H + \sqrt{\lambda_H}\beta_H).$$

The following proposition characterizes the first-best levels of the standardized excess capacities, denoted by β^*, and the corresponding first-best abandonment probabilities α^*; its proof is given in Appendix A.

Proposition 1 *Given* λ, *the first-best standardized excess capacity* β^* *is given by the unique solution of the following:*

$$\left[m(\beta^*)(1 - \frac{\Delta(\beta^*)}{\sqrt{\lambda}})\right]' = \frac{k}{\sqrt{\lambda}}, \tag{5}$$

and the corresponding first-best abandonment probability α^* is given by

$$\alpha^* = \frac{\Delta(\beta^*)}{\sqrt{\lambda}}. \tag{6}$$

Moreover, β^* is increasing and α^* is decreasing in λ.

Observe that at the first-best solution we achieve an efficient allocation in the sense that marginal benefits of excess capacity is equal to the marginal cost of staffing. Moreover, under the efficient allocation, the service provider supplies more excess capacity when the originator has a higher arrival rate. The originator's callers thus receive better service and have a lower abandonment probability. Note that the system profit must then be increasing in λ. Not only are more customers calling the system, they are on average less likely to abandon and will also spend more.

The proposition does not specify the payments between the parties. Once β^* is set, the overall profit of the system is fixed and the per-call payment merely splits the pie between the originator and the service provider. The exact value of that payment will depend on the relative bargaining power of the players. Assuming outside options are set to zero, the provider would prefer that c be set as close as possible to $m(\beta^*)$, which would leave the originator just indifferent to hiring the provider. The originator would prefer that the transfer price be $k\left(1 + \beta^*/\sqrt{\lambda}\right)$, which just allows the provider to recover his staffing costs.

3 Screening with Pay-per-call Contracts

We now suppose that the originator is privately informed about her arrival rate. The arrival rate can take two values λ_H, λ_L with $\lambda_H > \lambda_L$. For $i = L, H$, let β_i^* and α_i^* denote, respectively, the optimal standardized excess capacity and optimal abandonment probability when the arrival rate is known to be λ_i. In what follows, we will refer to an originator with arrival rate λ_H as the high type, and an originator with arrival rate λ_L as the low type. While it is obvious to presume the originator knows her market better than the provider, we also assume that the originator cannot simply relay this information to the provider in a credible manner. That is, the originator cannot simply produce market surveys that confirm her arrival rate. Consequently, unless the originator can take some additional action to demonstrate her type, the provider has only her prior probability p that the originator is a high type.

Here, we consider the service provider offering a pay-per-call contract to the originator without knowing her type. That is, the uninformed party offers a menu of contracts to the informed party to distinguish, or *screen*, an originator with a high arrival rate from one with a low rate. By the revelation principle, the provider can without loss of generality restrict his search for the optimal terms of trade to contracts of the following form: The call center offers a pair (α_i, c_i) specifying the abandonment probability α_i and the per-call payment c_i for each type $i = L, H$. Thus, the service provider tries to pick the optimal menu of contracts so as to maximize her expected profits subject to individual rationality and incentive compatibility conditions. The precise mathematical formulation of the service provider's problem is as follows: Choose $\{(\alpha_L, c_L), (\alpha_H, c_H)\}$ so as to

$$\text{maximize } p[c_H\lambda_H(1-\alpha_H) - k(\lambda_H + \sqrt{\lambda_H}\beta(\alpha_H, \lambda_H))]$$
$$+(1-p)[c_L\lambda_L(1-\alpha_L) - k(\lambda_L + \sqrt{\lambda_L}\beta(\alpha_L, \lambda_L))]$$

subject to

$$\lambda_H(1-\alpha_H)[m(\beta(\alpha_H, \lambda_H)) - c_H] \geq 0, \tag{IR$_H$}$$

$$\lambda_L(1-\alpha_L)[m(\beta(\alpha_L, \lambda_L)) - c_L] \geq 0, \tag{IR$_L$}$$

$$\lambda_H(1-\alpha_H)[m(\beta(\alpha_H, \lambda_H)) - c_H] \geq \lambda_H(1-\alpha_L)[m(\beta(\alpha_L, \lambda_H)) - c_L], \tag{IC$_H$}$$

$$\lambda_L(1-\alpha_L)[m(\beta(\alpha_L, \lambda_L)) - c_L] \geq \lambda_L(1-\alpha_H)[m(\beta(\alpha_H, \lambda_L)) - c_H]. \tag{IC$_L$}$$

Commonly referred to as participation constraints, the first two constraints impose individual rationality; they ensure that each type of originator prefers the contract designed to not hiring the provider[5]. The incentive compatibility constraint (IC$_i$) ensures that the originator of type i ($i = L, H$) prefers the contract devised for her over the contract devised for the other type. When both individual rationality and incentive compatibility constraints are satisfied, each type of originator will self select the contract devised for her. Note that the incentive compatibility constraints capture an important implicit assumption. The parties have contracted on an abandonment probability, not a staffing or excess capacity level. Hence, if a low-type originator were to take a contract designed for a high-type, the service provider must use excess capacity $\beta(\alpha_H, \lambda_L)$ which is greater than $\beta(\alpha_H, \lambda_H)$. This assumes that the service provider can quickly deduce the true arrival rate and adjust his staffing to fulfill his contractual obligation.

The profit of the service provider when serving the type i originator is equal to the pay-per-call

[5] We assume that the outside option of each type of originator is zero. However, the results are robust to introducing outside option K_i for type i provided $K_H/\lambda_H \geq K_L/\lambda_L$. A justification for assuming $K_H/\lambda_H \geq K_L/\lambda_L$ is that if the originator were to establish her own call center and incur the fixed setup plus variable staffing costs, the profit of the high type normalized by the arrival rate would be higher than that of the low type due to economies of scale in queueing systems.

fees $c_i\lambda_i(1-\alpha_i)$ minus the staffing cost necessary to support an abandonment probability of α_i. Therefore, the objective function of the service provider maximizes expected profits, where the expectation is taken with respect to the prior the service provider has on the originator's type. The following proposition characterizes the optimum menu of contracts and its proof is given in Appendix A.

Proposition 2 *The optimal menu of contracts, denoted by* $\{(\alpha_L, c_L), (\alpha_H, c_H)\}$, *offered by the service provider is unique and satisfies the following:*

$$\alpha_H < \alpha_H^* \quad \text{and } \alpha_L = \alpha_L^*,$$

where α_L^* *and* α_H^* *are the first-best abandonment probabilities. Moreover, the pay-per-call payments* c_L *and* c_H *are given as follows:*

$$c_H = m(\beta(\alpha_H, \lambda_H)) \quad \text{and } c_L = m(\beta(\alpha_L^*, \lambda_L)) + \frac{(1-\alpha_H)}{(1-\alpha_L^*)}[m(\beta(\alpha_H, \lambda_H)) - m(\beta(\alpha_H, \lambda_L))],$$

so that the high type originator earns zero profit whereas the low type gets a strictly positive information rent.

Relative to the first best, the optimal screening contract does not distort the abandonment probability α_L proffered to the low type but does alter the corresponding service level for the high type, offering an abandonment probability that is inefficiently low. Such distortion is unusual in the literature on monopolist screening. More typically, one has "efficiency at top" and the offering for the more favorable type (i.e., the type that is more profitable under full information) is not distorted and the favorable type earns an information rent while the less favorable type is pushed to indifference. The reversal of this standard result is driven by the statistical economies of scale phenomenon inherent in queueing systems. Economies of scale induce the service provider to distort the abandonment probability of the high type in order to reduce the information rent of the low type.

To see why this is necessary, notice that for any given α, we have $\beta(\alpha, \lambda_H) \leq \beta(\alpha, \lambda_L)$. In words, the service provider needs more standardized excess capacity to provide the low type with the same abandonment rate as the high type. Since the revenue (or margin) per answered call $m(\beta)$ earned by the originator is an increasing function of β, a low type originator who pretends to be a high type obtains a higher per call margin than a true high type originator would. Thus the first-best clearly cannot be implemented by pay-per-call contracts since a low volume originator would have an incentive to deviate and enjoy the benefits of enhanced service. However, it follows from Assumption 2 that the gain a low type captures by pretending to be a high type decreases as the abandonment rate under the high-type contract is reduced. Thus, if a boost in the high-type service level increases a high type's willingness to pay by a dollar, the low type's willingness to pay for the high service level increase by less than a dollar. Deviating to the high type's contract is consequently less attractive and the service provider can charge the low type a higher price (although the low type still earns a positive profit).

4 Signaling with Pay-per-call Contracts

We now reverse the order of play and have the originator offer a contract. There are two basic scenarios to consider. In the first, the high and low type originators offer distinct terms of trade. If the high type can devise a contract that the low is unwilling to copy, she effectively signals her private information to service provider. Alternatively, the two types of originators can pool, offering the same contract and leaving the service provider unable to garner additional information.

Recall that the service provider initially has prior beliefs that assign probability p to the originator being a high type and this is commonly known to all parties. Since the contract terms (c, α) may be informative, the service provider updates his beliefs about the type of the originator after seeing the contract offer. Let $\mu(c, \alpha)$ denote the belief of the service provider that the originator is of high type after observing a contract offer of (c, α). ($\mu(c, \alpha) = p$ if both types always offer the same contract.) The service provider makes the decision to accept or reject the contract based on his posterior $\mu(c, \alpha)$. In particular, he accepts the contract (c, α) if the expected payoff of accepting the contract exceeds zero.

We first consider the scenario in which the originator signals her information with the goal of characterizing the perfect Bayesian equilibria of this game. In our setting, a perfect Bayesian equilibrium corresponds to a set of strategies for the originator and the service provider, and a belief function μ of the service provider, which jointly satisfy the following:

(i) The strategies are optimal given the belief function μ.

(ii) The belief function μ is derived from the strategies through Bayes' rule whenever possible.

We restrict attention to pure strategy equilibria and focus on equilibria satisfying the intuitive criterion of Cho & Kreps (1987), consistent with most of the literature on signalling games; see, for instance Bagwell & Riordan (1991), Bagwell & Bernheim (1996), Schultz (1996) and Choi (1998).

The intuitive criterion is an equilibrium refinement which restricts beliefs off the equilibrium path. In particular, it requires that the updating of beliefs should not assign positive probability to a player taking an action that is equilibrium dominated (in a sense made precise in Appendix B.) Essentially, the intuitive criterion allows us to eliminate any perfect Bayesian equilibrium from which some type of originator would want to deviate even if she were not sure what exact belief the service provider would have as long as she knows that the provider would not think she is a type who would find the deviation equilibrium dominated. Appendix B reviews the formal definition of intuitive criterion and proves the following proposition (and others for this section), which provides an equivalent criterion in terms of the primitives of our problem. The following notation is needed to state Proposition 3. Let $\pi_i(c, \alpha)$ for $i = H, L$ denote the profit of the service provider who accepts contract (c, α) when the arrival rate is λ_i. That is,

$$\pi_i(c, \alpha) = (1 - \alpha)\lambda_i[m(\Delta^{-1}(\alpha\sqrt{\lambda_i})) - c].$$

Proposition 3 *A perfect Bayesian equilibrium with contract offers (α_L, c_L) and (α_H, c_H) violate the intuitive criterion of Cho & Kreps (1987), if and only if there exists a type $i \in \{L, H\}$ and a deviation contract (α, c) such that for $j \neq i$,*

$$\pi_i(\alpha, c) > \pi_i(\alpha_i, c_i),\ \pi_j(\alpha_j, c_j) > \pi_j(\alpha, c) \text{ and } (1 - \alpha)c\lambda_i - k(\lambda_i + \sqrt{\lambda_i}\Delta^{-1}(\alpha\sqrt{\lambda_i})) \geq 0. \tag{7}$$

Let (c_H, α_H) and (c_L, α_L) denote the contracts offered by the high and low type originators in a separating equilibria. The next proposition characterizes the separating equilibria under the intuitive criterion and is proved in Appendix B.

Proposition 4 *There exists a unique (pure strategy) perfect Bayesian equilibrium under the intuitive criterion which has the following properties*

$$\alpha_L = \alpha_L^* \text{ and } (1 - \alpha_L)c_L = k(1 + \frac{\Delta^{-1}(\alpha_L\sqrt{\lambda_L})}{\sqrt{\lambda_L}}),$$

$$\alpha_H < \alpha_H^* \text{ and } (1 - \alpha_H)c_H = k(1 + \frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}}).$$

Moreover, the low type is indifferent between her own contract and the contract offer of the high type, i.e. $\pi_L(\alpha_H, c_H) = \pi_L(\alpha_L, c_L)$.

Note that $\alpha_H < \alpha_H^*$. Just as the screening service provider in the previous section found it beneficial to distort the service level offered the high type, here the high type voluntarily decreases her abandonment probability below the first best. There are two consequence to this. First, she must pay more for every call; the high type now pays

$$(1-\alpha_H)c_H = k(1+\frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}})$$

per answered call. Second, she earns more for each call since $m(\beta)$ is increasing. Moreover, the originator is able to extract all the surplus from the service provider since she is making a take-it-or-leave-it offer. By Assumption 2, that gain in margin is less for the low type than for the high. Hence, while the high type lowers her profit (relative to the full information case) by asking for a higher service level, it makes mimicking her actions less attractive to the low type.

Before closing this section, we consider whether pooling equilibria might arise. In a pooling equilibrium, the high and low types offer the same contract (c, α) and the provider is unable to update his beliefs (i.e., $\mu(c, \alpha) = p$). The next proposition shows that a pooling equilibrium is *not* a possible outcome of the signalling game with reasonable beliefs; it is proved in Appendix B.

Proposition 5 *There exists no pooling equilibrium that satisfies the intuitive criterion.*

The intuition behind Proposition 5 is that the high type can always exploit economies of scale in order to distinguish herself from the low type while such a deviation would be dominated for the low type.

5 Introducing a Fixed Fee

We now expand the contracts the parties may use by introducing a fixed fee. The terms of trade are (α, c, T), where α and c are, as before, the agreed upon abandonment probability and per-call fee and T is a payment from the originator to the service provider that is independent of the realized call volume. There are several reasons for considering such a payment. First, call center outsourcing contracts frequently include such payments to cover initial training and set up costs. Second, they allow the contract terms to somewhat mimic the economies of scale of the underlying queuing system. Finally, two part tariffs have proven effective instruments in other studies of contracting under asymmetric information (see, for example, Chu (1992)). To see why they might be useful in our setting, note that for a given α, a high-type originator is indifferent between paying $c > 0$ per call answered with no fixed fee and paying a fixed fee of $c\lambda_H$ with no per-call charge. A low-type originator, however, obviously is not, preferring the low fixed payment and higher variable rate.

We now show a fixed fee may be sufficient to recover efficiency in both the screening and signalling scenarios. Proposition 6 deals with the former case while Proposition 7 deals with the latter.[6] Their proof are in Appendix C.

Proposition 6 *Let α_L^* and α_H^* denote the first-best abandonment probabilities. The optimal menu of contracts, denoted by $\{(\alpha_L, c_L, T_L), (\alpha_H, c_H, T_H)\}$, offered by the service provider is given as follows:*

i) If $\lambda_H m(\beta(\alpha_H^*, \lambda_H)) \geq \lambda_L m(\beta(\alpha_H^*, \lambda_L))$, then the service provider offers the first-best abandonment probabilities. That is,

$$\alpha_H = \alpha_H^* \quad \text{and } \alpha_L = \alpha_L^*. \tag{8}$$

As for the per-call payments, the service provider can choose any (c_L, c_H) such that

$$c_H \leq \frac{\lambda_H m(\beta(\alpha_H, \lambda_H)) - \lambda_L m(\beta(\alpha_H, \lambda_L))}{\lambda_H - \lambda_L}, \tag{9}$$

$$c_L \geq \frac{\lambda_H m(\beta(\alpha_L, \lambda_H)) - \lambda_L m(\beta(\alpha_L, \lambda_L))}{\lambda_H - \lambda_L}, \tag{10}$$

provided that the fixed fees T_L and T_H are as follows:

$$T_H = \lambda_H(1-\alpha_H)m(\beta(\alpha_H, \lambda_H)) - \lambda_H(1-\alpha_H)c_H, \tag{11}$$

$$T_L = \lambda_L(1-\alpha_L)m(\beta(\alpha_L, \lambda_L)) - \lambda_L(1-\alpha_L)c_L. \tag{12}$$

ii) If $\lambda_H m(\beta(\alpha_H^*, \lambda_H)) < \lambda_L m(\beta(\alpha_H^*, \lambda_L))$, then the abandonment probabilities offered by the service provider are uniquely determined. To be specific, the service provider offers the first-best abandonment probability to the low type, i.e. $\alpha_L = \alpha_L^*$, while he offers a lower abandonment probability to the high type than the first-best level, i.e. $\alpha_H < \alpha_H^*$. Moreover, the fixed fee and per-call payment of the high type are uniquely determined as follows:

$$c_H = 0 \quad \text{and} \quad T_H = \lambda_H (1-\alpha_H) m (\beta(\alpha_H, \lambda_H)),$$

so that the high type originator earns zero profits, while the service provider can choose any per-call payment c_L for the low type such that

$$\begin{aligned} c_L \quad \geq \quad & \frac{\lambda_H[m(\beta(\alpha_L, \lambda_H))] - \lambda_L[m(\beta(\alpha_L, \lambda_L))]}{\lambda_H - \lambda_L} \\ & + \frac{(1-\alpha_H)}{(1-\alpha_L)}\left[\frac{\lambda_L}{\lambda_H - \lambda_L} m(\beta(\alpha_H, \lambda_L)) - \frac{\lambda_H}{\lambda_H - \lambda_L} m(\beta(\alpha_H, \lambda_H))\right], \end{aligned} \tag{13}$$

[6] We do not explicitly consider pooling when the originator offers the contract. One can derive a result similar to Proposition 5 that rules out a pooling equilibrium.

provided

$$T_L = \lambda_L (1-\alpha_L) [m(\beta(\alpha_L,\lambda_L)) - c_L] + (1-\alpha_H) [\lambda_H m(\beta(\alpha_H,\lambda_H)) - \lambda_L m(\beta(\alpha_H,\lambda_L))]. \quad (14)$$

In particular,

$$T_L + \lambda_L (1-\alpha_L) c_L = \lambda_L (1-\alpha_L) m(\beta(\alpha_L,\lambda_L)) + (1-\alpha_H) [\lambda_H m(\beta(\alpha_H,\lambda_H)) - \lambda_L m(\beta(\alpha_H,\lambda_L))],$$

so that the low type originator earns positive information rent.

Proposition 7 *Let α_L^* and α_H^* denote the first-best abandonment probabilities.*

(i) If

$$\begin{aligned} & (1-\alpha_L^*) \lambda_L m(\beta(\alpha_L^*,\lambda_L)) - k(\lambda_L + \sqrt{\lambda_L}\Delta^{-1}(\alpha_L^*\sqrt{\lambda_L})) \\ \geq \; & (1-\alpha_H^*) \lambda_L m(\beta(\alpha_H^*,\lambda_L)) - k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha_H^*\sqrt{\lambda_H})), \end{aligned} \quad (15)$$

then any (pure strategy) separating perfect Bayesian equilibrium under the intuitive criterion has the following properties:

$$\alpha_L = \alpha_L^* \quad \text{and} \quad \alpha_H = \alpha_H^*,$$

that is, the low and high type originators offer the first-best abandonment probabilities. As for the per-call payments, the originators can choose any (c_L, c_H) such that

$$\begin{aligned} c_H \leq \; & \frac{(1-\alpha_L)\lambda_L m(\beta(\alpha_L,\lambda_L)) - (1-\alpha_H)\lambda_L m(\beta(\alpha_H,\lambda_L))}{(1-\alpha_H)(\lambda_H-\lambda_L)} \\ & + \frac{k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha_H\sqrt{\lambda_H}) - k(\lambda_L + \sqrt{\lambda_L}\Delta^{-1}(\alpha_L\sqrt{\lambda_L}))}{(1-\alpha_H)(\lambda_H-\lambda_L)} \end{aligned} \quad (16)$$

$$\begin{aligned} c_L \geq \; & \frac{(1-\alpha_L)\lambda_H m(\beta(\alpha_L,\lambda_H)) - (1-\alpha_H)\lambda_H m(\beta(\alpha_H,\lambda_H))}{(1-\alpha_L)(\lambda_H-\lambda_L)} \\ & + \frac{k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha_H\sqrt{\lambda_H}) - k(\lambda_L + \sqrt{\lambda_L}\Delta^{-1}(\alpha_L\sqrt{\lambda_L}))}{(1-\alpha_L)(\lambda_H-\lambda_L)}, \end{aligned} \quad (17)$$

provided that the fixed fees T_L and T_H are as follows:

$$\begin{aligned} T_H &= k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha_H\sqrt{\lambda_H})) - \lambda_H(1-\alpha_H)c_H, \quad (18) \\ T_L &= k(\lambda_L + \sqrt{\lambda_L}\Delta^{-1}(\alpha_L\sqrt{\lambda_L})) - \lambda_L(1-\alpha_L)c_L. \quad (19) \end{aligned}$$

so that both the high and low type originator extract all the surplus.

(ii) If (15) does not hold, then for any (pure strategy) separating perfect Bayesian equilibrium under the intuitive criterion, the abandonment probabilities are determined uniquely, and they satisfy the following:

$$\alpha_L = \alpha_L^* \quad \text{and} \quad \alpha_H < \alpha_H^*,$$

where α_H is such that the low type originator is indifferent between her contract and the contract offer of the high type; and the low type originator offers the first-best abandonment probability. Moreover, the fixed fee and per-call payment of the high type originator, c_H and T_H, are also uniquely determined as follows

$$c_H = 0,\, T_H = k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha_H\sqrt{\lambda_H})),$$

whereas the low type originator can choose any per-call payment c_L such that

$$\begin{aligned} c_L \quad \geq \quad & \frac{(1-\alpha_L)\,\lambda_H m\,(\beta\,(\alpha_L,\lambda_H)) - (1-\alpha_H)\,\lambda_H m\,(\beta\,(\alpha_H,\lambda_H))}{(1-\alpha_L)\,(\lambda_H-\lambda_L)} \\ & + \frac{k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha_H\sqrt{\lambda_H}) - k(\lambda_L + \sqrt{\lambda_L}\Delta^{-1}(\alpha_L\sqrt{\lambda_L}))}{(1-\alpha_L)\,(\lambda_H-\lambda_L)}, \end{aligned} \tag{20}$$

provided

$$T_L = k(\lambda_L + \sqrt{\lambda_L}\Delta^{-1}(\alpha_L\sqrt{\lambda_L})) - \lambda_L(1-\alpha_L)c_L, \tag{21}$$

and, hence there is multiplicity in the choice of (c_L, T_L).

Adding a fixed fee has similar affects on both the screening and signalling scenarios. For both, if there is sufficient differentiation between the volume of calls for high and low types (exactly what is sufficient depends on the scenario), a fixed fee is enough to restore efficiency. Both types receive their first best service levels, and (in the screening case) the low type no longer captures any information rents. In these setting multiple possible equilibrium contracts exist. Intuitively, one option is always to have the high type pay just a fixed fee with no per-call charge. If the low type strictly prefers her full information contract to these terms, there is a multiplicity of contract because the fixed fee of the high type could be lowered and a per-call charge added that leaves the total payment of the high type unchanged but is still unattractive to the low type.

This changes when there is not much difference between the arrival rates. Now if the high type were to receive its first best service level and pay only a fixed fee, a low type would want to take that contract. Effective screening or signalling thus again requires distorting the service level. The amount of the distortion is less than before. By sticking with a fixed fee and no per-call charge, the equilibrium takes the contract form that is least attractive to the low type in order to minimize the system loss due to deviation from the first best. In the case of screening, this also lowers the information rent paid to the low type.

Note that the fixed fee contract has much in common with the terms offered by call center vendors. As discussed above, Accolade Support offers packages that require originators to essentially pre-pay for a fixed number of minutes and then pay higher per minute rates when demand exceeds the pre-paid quantity. Similarly, Centris Information Systems may set a minimum utilization which effectively commits the originator to paying for some number of calls. Here, under both screening and signalling, it is always possible to have high volume originator paying just a fixed fee with no per-call charge. This is equivalent to committing to a minimum purchase quantity that a low volume originator finds unattractive.

6 Discussion

We have examined how asymmetric demand information can affect contracting between two parties. Where past studies have focused on deterministic demand curves or newsvendor problems, we assume a call center, an important and growing part of the business process outsourcing market. Focusing on a call center naturally leads us to using a queuing framework. Queuing systems exhibit economies of scale and this leads to significant changes in the contract terms. In particular, while monopoly screening problems do not generally recommend distorting the offering to the most favorable type, here we find that it is optimal to offer a high volume originator an inefficiently high level of service. Further, a low volume originator ends up earning a positive information rent. One sees a similar distortion when we switch to a signalling setting; a high type requests (and pays for) a service level so high that the low type does not find it worth copying her request. The amount of distortion in both the screening and signalling settings is reduced if the contracts are expanded to include a fixed fee.

We have taken the abandonment probability as our measure of service. This is largely done to increase the transparency of the presentation. Alternatively, one could have other service level criterion such as an upper bound on the probability of waiting more than a certain amount. Indeed, for such criterion, one can use approximations similar to ours to show that our structural results continue to hold.

A question remains: What determines the extent of the distortion from the first best? As one might expect, there is less distortion in the screening scenario relative to the first-best as the proportion of high volume originators increase. If the proportion of high volume originators is high, then the cost associated with distorting the allocation of the high volume operator outweighs the benefit of distinguishing (screening) different types of originators. On the other hand, in the signalling scenario, the actions of the low and high volume operators are not affected by the proportion of the high types in the population.

Another important model element is the revenue function $m(\cdot)$. In particular, the curvature of m impacts the distortion in the service levels. To investigate the effect of the curvature of the revenue function $m(\cdot)$, we conduct a numerical example with the following data: We set the arrival rates as $\lambda_L = 40$, $\lambda_H = 100$ and the abandonment rate as $\eta = 1$. Finally, we have $k = 6$ and $p = 0.25$. To isolate the effect of the curvature of $m(\cdot)$, we proceed as follows. First consider the contracting problem for the following function, which constitutes our base case: $m(\beta) = 8 - 1/\beta^{0.1}$, for which the first best excess capacity levels are

$$\beta_H^* = 0.32 \text{ and } \beta_L^* = 0.22.$$

To isolate the effect of the curvature on the distortions in the optimal contracts, we consider revenue functions m_n for $n \geq 0.1$, where we only change the function to the right of β_H^* as n changes. We also make sure that m_n is smooth. This construction ensures that the first-best excess capacity levels are the same for each n. Our construction also ensures that the revenue function m_n becomes more an more flat to the right of β_H^* as n increases. To be specific, the functions we consider are given as follows: For $n \geq 0.1$

$$m_n(\beta) = \begin{cases} 8 - 1/\beta^{0.1} & \text{if } \beta \leq \beta_H^* \\ w_n - s_n/\beta^n & \text{if } \beta > \beta_H^* \end{cases},$$

where $w_n = 8 + (0.1/n - 1)/(\beta_H^*)^{0.1}$ and $s_n = (0.1/n)(\beta_H^*)^{n-0.1}$. Figure 1 depicts the revenue function for different values of n. The curvature of m increases with n.

The tables we provide next display how the distortion of the abandonment probability α_H of the high volume originator relative to the first-best level α_H^* changes in the screening and signalling scenarios as the curvature of the revenue function increases.

In both cases the distortions decreases as n increases. Intuitively, as the revenue function becomes flatter, it becomes more costly to separate the high and low volume originators using a distortion of standardized excess capacity.

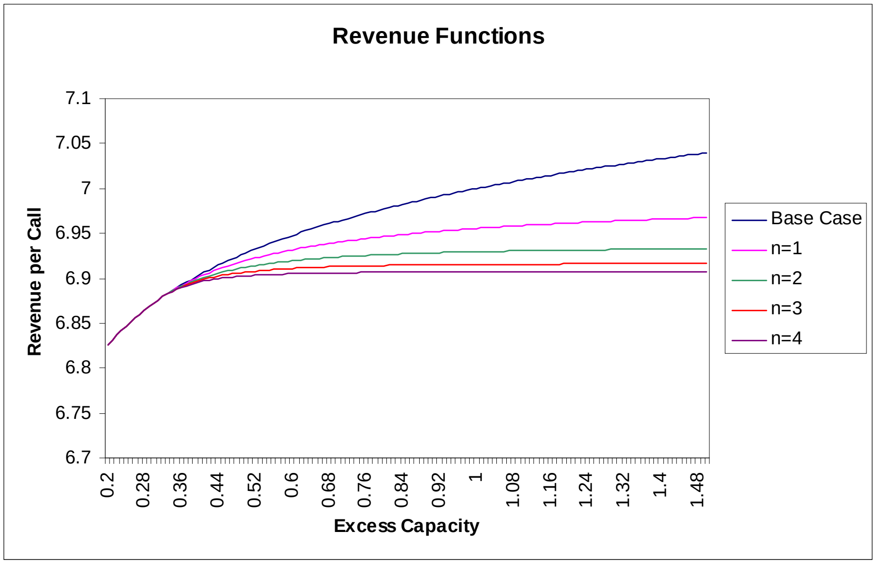

Figure 1: Revenue functions for different values of n.

n	β_H^*	β_L^*	α_H^*	α_L^*	Screening β_H	Screening α_H	% change in α	$(\alpha_H^* - \alpha_H)$
0.1	0.32	0.22	0.026	0.047	0.48	0.0205	21.1%	0.0055
1	0.32	0.22	0.026	0.047	0.44	0.0217	16.5%	0.0043
2	0.32	0.22	0.026	0.047	0.41	0.0226	13.1%	0.0034
3	0.32	0.22	0.026	0.047	0.39	0.0233	10.4%	0.0027
4	0.32	0.22	0.026	0.047	0.38	0.0237	8.8%	0.0023

Table 1: Impact of the curvature of the revenue function on the optimal contract in screening case.

n	β_H^*	β_L^*	α_H^*	α_L^*	Signalling β_H	Signalling α_H	% change in α	$(\alpha_H^* - \alpha_H)$
0.1	0.32	0.22	0.026	0.047	1.35	0.006	76.9%	0.020
1	0.32	0.22	0.026	0.047	1.23	0.008	69.2%	0.018
2	0.32	0.22	0.026	0.047	1.17	0.009	65.4%	0.017
3	0.32	0.22	0.026	0.047	1.14	0.010	61.5%	0.016
4	0.32	0.22	0.026	0.047	1.12	0.011	57.7%	0.015

Table 2: Impact of the curvature of the revenue function on the optimal contract: in signalling case.

References

Aksin, O., de Vericourt, F. & Karaesmen, F. (2006), 'Call center outsourcing contract design and choice', *Management Science* . To appear.

Bagwell, K. & Bernheim, B. (1996), 'Venlen effects in a theory of conspicuous consumption', *American Economic Review* **86**.

Bagwell, K. & Riordan, M. (1991), 'High and declining prices signal product quality', *American Economic Review* **81**.

Beasty, C. (2005), 'The call center outsourcing outlook.', *Business Insights* .

Beasty, C. (2006), 'Contact center outsourcing strengthens', *Destination CRM.com* p. January 17.

Cachon, G. (2003), *Supply chain coordination with contracts*, Handbooks in Operations Research and Management Science: Supply Chain Management.

Cachon, G. & Harker, P. (2002), 'Competition and outsourcing with scale ecomonies', *Management Science* **48**(10), 1314–1333.

Cachon, G. & Lariviere, M. P. (2001), 'Contracting to assure supply: How to share demand forecasts in a supply chain', *Management Science* **47**(5), 629–646.

Cho, I. & Kreps, D. (1987), 'Signalling games and stable equilibria', *Quarterly Journal of Economics* **102**, 179–221.

Choi, J. (1998), 'Brand extension as informatioal leverage', *Review of Economic Studies* **65**.

Chu, W. (1992), 'Demand signaling and screening in channels of distribution', *Marketing Science* **11**(4), 327–347.

Cohen, L. & Young, A. (2006), *Moving beyond outsourcing to Achieve Growth and Agility*, Harvard Business School Press.

Desai, P. & Sirinivasan, K. (1995), 'A franchise management issue:Demand signaling under unobservable service', *Management Science* **41**(10), 1608–1623.

Garnett, O., Mandelbaum, A. & Reiman, M. (2002), 'Designing a call center with impatient customers', *Manufacturing & Service Operations Management* **4**, 208–227.

Hasija, S., Pinker, A. & Shumsky, R. (2006), 'Call center outsourcing contracts under asymmetric information'. Working Paper.

Hassin, R. & Haviv, M. (2002), *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems*, Kluwer Academic Publishers.

Milner, J. & Olsen, T. (2007), 'Service level agreements in call centers: Perils and prescriptions', *Management Science* . To appear.

Munoz, C. (2007), 'External affairs.', *The Economist* **384**, 65–66.

Niezen, C. & Weller, W. (2006), 'Procurement as strategy', *Harvard Business Review* **84**(9), 22–24.

Ren, Z. & Zhang, F. (2007), 'Service outsourcing: Information asymmetry and service quality'. Working paper.

Ren, Z. & Zhou, Y.-P. (2006), 'Call center outsourcing: Coordinating staffing levels and service quality', *Management Science* . Forthcoming.

Salanié, B. (1997), *The Economics of Contracts*, The MIT Press.

Schultz, C. (1996), 'Polarization and inefficient policies', *Review of Economic Studies* **63**.

Taylor, R. & Tofts, C. (2006), 'Death by a thousand slas; A short study of commerical suicide pacts.'. working paper. HP Laboratories Bristol.

A Proofs in Sections 2.1 and 3

Proof of Proposition 1. The first order condition, which is necessary and sufficient by Assumption 1, gives

$$\left[m(\beta^*)(1-\frac{\Delta(\beta^*)}{\sqrt{\lambda}})\right]' = k/\sqrt{\lambda}.$$

By Assumption 1, we have β^* *is* increasing in λ. Then, α^* is decreasing since Δ is decreasing. ■

Lemma 1 *The constraints* (IR_H) *and* (IC_L) *of the service provider's screening problem bind at an optimum contract.*

Proof of Lemma 1. Suppose (IR_H) does not bind. Then, increasing c_H by a small amount is more profitable to the service provider. Clearly, we can do this if (IC_H) does not bind, because increasing c_H relaxes (IC_L). Therefore, we restrict attention to the possibility that (IC_H) binds. In that case, we have that

$$(1-\alpha_H)[m(\beta(\alpha_H,\lambda_H))-c_H] = (1-\alpha_L)[m(\beta(\alpha_L,\lambda_H))-c_L]. \tag{22}$$

Since (IR_H) does not bind by assumption, we have $m(\beta(\alpha_H,\lambda_H)) > c_H$. Then, (22) implies that $m(\beta(\alpha_L,\lambda_H)) > c_L$ as well. Because c_L and c_H are related through (22), we need to check that

(IR_L) is not violated as we increase c_H by a small amount. Thus, we next consider (IR_L). Since $m(\cdot)$ is increasing and $\beta(\alpha_L, \lambda_L)) > \beta(\alpha_L, \lambda_H)$, we have

$$m(\beta(\alpha_L, \lambda_L)) - c_L \geq m(\beta(\alpha_L, \lambda_H)) - c_L > 0.$$

Therefore, (IR_L) does not bind either. Then, consider increasing c_H by $\frac{\epsilon}{1-\alpha_H}$ and c_L by $\frac{\epsilon}{1-\alpha_L}$ for sufficiently small $\epsilon > 0$. Then, the incentive compatibility constraints are not violated since we decrease both sides by ϵ, but the objective value of the service provider's screening problem increases strictly, which contradicts optimality. Thus, (IR_H) binds at an optimum contract.

Now suppose that (IC_L) does not bind. Since (IR_H) binds by the above argument, we have that $m(\beta(\alpha_H, \lambda_H)) = c_H$. Then, IC_L becomes

$$(1-\alpha_L)[m(\beta(\alpha_L, \lambda_L)) - c_L] > (1-\alpha_H)[m(\beta(\alpha_H, \lambda_L)) - m(\beta(\alpha_H, \lambda_H))],$$

where the right hand side is strictly positive since $\beta(\alpha_H, \lambda_L) > \beta(\alpha_H, \lambda_H)$, which implies that ($\text{IR}_L$) does not bind either. Also note that increasing c_L by a sufficiently small $\epsilon > 0$ relaxes (IC_H). Thus, we can increase c_L by a sufficiently small $\epsilon > 0$, in which case (IR_L) and (IC_L) are still satisfied and we get a strictly better objective value, contradicting the optimality. Thus, (IC_L) binds. ■

Proof of Proposition 2. First notice that we have the following relations for any given abandonment probabilities α_L and α_H:

$$\begin{aligned} \beta(\alpha_L, \lambda_H) &\leq \beta(\alpha_L, \lambda_L), \\ \beta(\alpha_H, \lambda_L) &\geq \beta(\alpha_H, \lambda_H), \end{aligned}$$

which is due to statistical economies of scale inherent in queueing systems. Also recall that (IR_H) and (IC_L) binds at an optimum contract by Lemma 1. Then, (IR_L) is also satisfied since

$$\begin{aligned} \lambda_L(1-\alpha_L)[m(\beta(\alpha_L, \lambda_L)) - c_L] &= \lambda_L(1-\alpha_H)[m(\beta(\alpha_H, \lambda_L)) - c_H], \\ &\geq \lambda_L(1-\alpha_H)[m(\beta(\alpha_H, \lambda_H)) - c_H], \\ &= 0, \end{aligned}$$

where the first equality follows from the fact that (IC_L) binds, the second line is true since $\beta(\alpha_H, \lambda_L) \geq \beta(\alpha_H, \lambda_H)$ and the last line follows from (IR_H).

We proceed with the service provider's screening problem ignoring the constraint (IC_H) for now. We will first characterize the optimal solution to a relaxed problem, which ignores (IC_H), and then verify that (IC_H) is indeed satisfied by that solution in the end. Given that (IR_H) and IC_L bind by Lemma 1, we can solve for c_H and c_L, which gives the following:

$$c_H = m(\beta(\alpha_H, \lambda_H)), \tag{23}$$

$$c_L = m(\beta(\alpha_L, \lambda_L)) + \frac{(1-\alpha_H)}{(1-\alpha_L)}[m(\beta(\alpha_H, \lambda_H)) - m(\beta(\alpha_H, \lambda_L))]. \tag{24}$$

Then, substituting (23) and (24) into the objective function of the service provider's problem (ignoring (IC_H) for now) we arrive at the following problems. Choose the abandonment probabilities α_L and α_H so as to maximize

$$\begin{aligned} &p[m(\beta(\alpha_H, \lambda_H))\lambda_H(1-\alpha_H) - k(\lambda_H + \sqrt{\lambda_H}\beta(\alpha_H, \lambda_H))] - (1-p)k(\lambda_L + \sqrt{\lambda_L}\beta(\alpha_L, \lambda_L)) \\ &+(1-p)\left[\left(m(\beta(\alpha_L, \lambda_L)) + \frac{(1-\alpha_H)}{(1-\alpha_L)}[m(\beta(\alpha_H, \lambda_H)) - m(\beta(\alpha_H, \lambda_L))]\right)\lambda_L(1-\alpha_L)\right] \end{aligned} \tag{25}$$

It is easy to see that the optimal level of α_L is equal to α_L^*, the first-best level. To find the optimum α_H, we consider the following problem, which ignores the terms that do not depend on α_H: Choose the abandonment probability α_H for the high type so as to

$$\begin{aligned} &\text{maximize}_{\alpha_H}\ p[m(\beta(\alpha_H, \lambda_H))\lambda_H(1-\alpha_H) - k(\lambda_H + \sqrt{\lambda_H}\beta(\alpha_H, \lambda_H))] \\ &+(1-p)\left[((1-\alpha_H)[m(\beta(\alpha_H, \lambda_H)) - m(\beta(\alpha_H, \lambda_L))])\lambda_L\right]. \end{aligned} \tag{26}$$

By (3), the maximization problem (26) can equivalently be stated as a problem of choosing the optimal standardized excess capacity β_H as follows: Choose β_H so as to

$$\begin{aligned} &\text{maximize } p[m(\beta_H)\lambda_H\left(1 - \frac{\Delta(\beta_H)}{\sqrt{\lambda_H}}\right) - k(\lambda_H + \sqrt{\lambda_H}\beta_H)] \\ &+(1-p)\lambda_L\left[\left(1 - \frac{\Delta(\beta_H)}{\sqrt{\lambda_H}}\right)[m(\beta_H) - m(\widehat{\beta}_H)]\right], \end{aligned} \tag{27}$$

where $\widehat{\beta}_H$ is defined as a function of β_H through the relation

$$\frac{\Delta(\beta_H)}{\sqrt{\lambda_H}} = \frac{\Delta(\widehat{\beta}_H)}{\sqrt{\lambda_L}}.$$

The first order conditions for (27) give

$$\left[m(\beta_H)\left(1 - \frac{\Delta(\beta_H)}{\sqrt{\lambda_H}}\right)\right]' = \frac{k}{\sqrt{\lambda_H}} - \frac{(1-p)\lambda_L}{p\lambda_H}\left[\left(1 - \frac{\Delta(\beta_H)}{\sqrt{\lambda_H}}\right)[m(\beta_H) - m(\widehat{\beta}_H)]\right]', \tag{28}$$

where the second term $\left[(1 - \frac{\Delta(\beta_H)}{\sqrt{\lambda_H}})[m(\beta_H) - m(\widehat{\beta}_H)]\right]' \geq 0$ by Assumption 2. To see this, notice that Assumption 2 can be rewritten as follows using simple algebraic manipulations:

$$(1-\alpha)[m(\Delta^{-1}(\alpha\sqrt{\lambda})) - m(\Delta^{-1}(\alpha\sqrt{\kappa\lambda}))] \text{ is decreasing as } (1-\alpha) \text{ increases for } \kappa > 1.$$

Then, the term $\left[(1-\frac{\Delta(\beta_H)}{\sqrt{\lambda_H}})[m(\beta_H)-m(\widehat{\beta}_H)]\right]$ can be expressed as a function of the abandonment probability α_H through the relation (3) as follows:

$$(1-\alpha_H)\left[m\left(\Delta^{-1}\left(\alpha_H\sqrt{\lambda_H}\right)\right)-m\left(\Delta^{-1}\left(\alpha_H\sqrt{\lambda_L}\right)\right)\right],$$

which is nonpositive and increasing as α_H decreases by Assumption 2. This is equivalent to have $\left[(1-\frac{\Delta(\beta_H)}{\sqrt{\lambda_H}})[m(\beta_H)-m(\widehat{\beta}_H)]\right]$ increasing as a function of β_H, cf. (3). Then comparing (5) with (28) yields $\beta_H \geq \beta_H^*$ by Assumption 1, which in turn implies, cf. (2), that

$$\alpha_H=\frac{\Delta\left(\beta_H\right)}{\sqrt{\lambda_H}}\leq\frac{\Delta\left(\beta_H^*\right)}{\sqrt{\lambda_H}}=\alpha_H^*<\alpha_L^*=\alpha_L.$$

Then it follows from Assumption 2 that

$$\begin{aligned}&(1-\alpha_L)\left[m\left(\Delta^{-1}\left(\alpha_L\sqrt{\lambda_L}\right)\right)-m\left(\Delta^{-1}\left(\alpha_L\sqrt{\lambda_H}\right)\right)\right]\\ \geq\ &(1-\alpha_H)\left[m\left(\Delta^{-1}\left(\alpha_H\sqrt{\lambda_L}\right)\right)-m\left(\Delta^{-1}\left(\alpha_H\sqrt{\lambda_H}\right)\right)\right],\end{aligned}$$

which in turn is equivalent to (cf. 3)

$$(1-\alpha_L)[m(\beta(\alpha_L,\lambda_L))-m(\beta(\alpha_L,\lambda_H))]\geq(1-\alpha_H)[m(\beta(\alpha_H,\lambda_L))-m(\beta(\alpha_H,\lambda_H))]. \tag{29}$$

Recall that we had ignored (IC$_H$) while we solve the above maximization problem. What remains is to check whether (IC$_H$) is satisfied in this solution. For that,the right hand side of (IC$_H$) is zero, since (IR$_H$) binds. Then rewrite (IC$_H$) as follows:

$$0\geq(1-\alpha_L)[m(\beta(\alpha_L,\lambda_H))-c_L]. \tag{30}$$

Substituting (24) into (30), an rearranging terms we see that checking whether (IC$_H$) is satisfied is equivalent to checking whether the following holds:

$$(1-\alpha_L)[m(\beta(\alpha_L,\lambda_L))-m(\beta(\alpha_L,\lambda_H))]\geq(1-\alpha_H)[m(\beta(\alpha_H,\lambda_L))-m(\beta(\alpha_H,\lambda_H))],$$

which is precisely (29). Thus, (IC$_H$) is satisfied.

Finally, note that the high type originator earns zero profits by (23); and it follows from (24) and the fact that $m(\beta(\alpha_H,\lambda_L))>m(\beta(\alpha_H,\lambda_H))$ that the low type originator gets strictly positive information rent. ■

B Proofs and Auxiliary Derivations for Section 4

In this appendix, we prove the results in Section 4 and provide some auxiliary results necessary for proofs. We also reviews the formal definition of the intuitive criterion of Cho & Kreps (1987), which we specialize for our setting.

Intuitive Criterion. The idea of this refinements is that reasonable beliefs should not assign positive probability to a player taking an action that is strictly dominated (in a sense to be made precise) for her. To formalize this, for any nonempty set $\Theta \subset \{H, L\}$, let $S^*(\Theta, (c, \alpha)) \subset \{\text{Accept}, \text{Reject}\}$ denote the set of possible equilibrium responses by the service provider that can arise after contract offer (c, α) is observed for *some* beliefs satisfying the property that $\mu(c, \alpha) = 1$ if $\Theta = \{H\}$ and $\mu(c, \alpha) = 0$ if $\Theta = \{L\}$. That is, the set $S^*(\Theta, (c, \alpha))$ contains the equilibrium responses by the service provider to the contract choice (c, α) for some beliefs that assign positive probability to types only in the set Θ. When we have $\Theta = \{H, L\}$, this construction puts no restriction on the beliefs.

We now introduce the notion of equilibrium dominance. To facilitate our analysis, we introduce the following definitions with a slight abuse of notation. Let $\pi_H(c, \alpha; s)$ denote the profit of a high type originator if she offers the contract (c, α) and the service provider's decision is $s \in \{\text{Accept}, \text{Reject}\}$. Then, we have

$$\pi_H(c, \alpha; \text{Accept}) = (1 - \alpha)\lambda_H[m(\Delta^{-1}(\alpha\sqrt{\lambda_H})) - c] \;\text{ and }\; \pi_H(c, \alpha; \text{Reject}) = 0.$$

Similarly define $\pi_L(c, \alpha; s)$ as the profit of the low type originator if she offers a contract (c, α) and the service provider's decision is $s \in \{\text{Accept, Reject}\}$. Let (c_H, α_H) and (c_L, α_L) denote the contract offers of the high and low type originators in a perfect Bayesian equilibrium with belief system μ. The equilibrium responses by the call center to the contract offers (c_H, α_H) and (c_L, α_L) should be to accept since otherwise the originator would not be maximizing her payoff. We then say that a contract (c, α) is equilibrium dominated for the high type in this perfect Bayesian equilibrium if

$$(1 - \alpha_H)\lambda_H[m(\Delta^{-1}(\alpha_H\sqrt{\lambda_H})) - c_H] \; > \text{Max}_{s \in S^*(\{H,L\},(c,\alpha))} \; \pi_H(c, \alpha; s). \tag{31}$$

We similarly define the contracts that are equilibrium dominated for the low type. Then, for each contact (c, α), we define the set $\Theta^*(c, \alpha)$ as the set of types for which (c, α) is not equilibrium dominated. Finally, the perfect Bayesian equilibrium with contract offers (c_H, α_H) and (c_L, α_L) and belief system μ is said to violate the intuitive criterion if there exists a type $\theta \in \{H, L\}$ and a contract (c, α) such that

$$\text{Min}_{s \in S^*(\Theta^*(c,\alpha),\ (c,\alpha))}\ \pi_\theta(c, \alpha; s) > \pi_\theta(c_\theta, \alpha_\theta; \text{Accept}). \tag{32}$$

For instance, a pooling equilibrium with contract offer $(\widetilde{c}, \widetilde{\alpha})$ and the belief system μ violate the intuitive criterion if there exists a type θ and a deviation (c, α) that yields higher profit for type θ than her equilibrium profit as long as the call center does not assign a positive probability to a type for which the deviation (c, α) is equilibrium dominated.

The following lemma is nothing but the "if part" of Proposition 3, which is really what is needed for the following proofs. Thus, we next state and prove Lemma 2; and we will eventually prove the "only if" part of Proposition 3, which provides a criterion (in terms of problem primitives) equivalent to the intuitive criterion in our setting.

Lemma 2 *A perfect Bayesian equilibrium with contract offers (c_L, α_L) an (c_H, α_H) violate the intuitive criterion of Cho & Kreps (1987) if there exists a type $i \in \{L, H\}$ and a deviation contract $(\widetilde{c}, \widetilde{\alpha})$ such that*

$$\pi_i(\widetilde{\alpha}_i, \widetilde{c}_i) > \pi_i(\alpha_i, c_i),\ \pi_j(\alpha_j, c_j) > \pi_j(\widetilde{\alpha}_i, \widetilde{c}_i) \text{ and } (1-\widetilde{\alpha}_i)\widetilde{c}_i\lambda_i - k(\lambda_i + \sqrt{\lambda_i}\Delta^{-1}(\widetilde{\alpha}_i\sqrt{\lambda_i})) \geq 0 \text{ for } j \neq i. \tag{33}$$

Proof of Lemma 2. In any equilibrium we have $\pi_i(c_i, \alpha_i;\text{Accept}) \geq 0$ for $i = L, H$. Thus, in (32), we should have $S^*(\Theta^*(c, \alpha),\ (c, \alpha)) = \{\text{Accept}\}$. This, in turn, implies that a contract (α, c) is a deviation for type $i = L, H$ violating the intuitive criterion if $\pi_i(\alpha, c) > \pi_i(\alpha_i, c_i)$, $\pi_j(\alpha_j, c_j) > \pi_j(\alpha, c)$ for $j \neq i$ (which guarantees that $\Theta^*(c, \alpha) = \{i\}$) and $(1-\alpha)c\lambda_i - k(\lambda_i + \sqrt{\lambda_i}\Delta^{-1}(\alpha\sqrt{\lambda_i})) \geq 0$ (which ensures $S^*(\Theta^*(c, \alpha),\ (c, \alpha)) = \{\text{Accept}\}$). ■

We will use the following lemma to construct deviations of the form given in Lemma 2; and we will eventually prove the "only if" part of Proposition 3, which provides a criterion at various places in what follows.

Lemma 3 *Consider candidate equilibrium contract offers (c_L, α_L) and (c_H, α_H) (not necessarily different, i.e. we allow for $(c_L, \alpha_L) = (c_H, \alpha_H)$) offered by the low and high type originators, respectively. If $\pi_H(c_H, \alpha_H) = \pi_H(c_L, \alpha_L)$, then there exists a deviation (α, c) for the high type such that*

$$\pi_H(\alpha, c) > \pi_H(\alpha_H, c_H),\ \pi_L(\alpha_L, c_L) > \pi_L(\alpha, c) \text{ and } (1-\alpha)c\lambda_H - k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha\sqrt{\lambda_H})) \geq 0.$$

Thus, $\{(c_L, \alpha_L), (c_H, \alpha_H)\}$ cannot correspond to outcomes of a perfect Bayesian equilibrium satisfying the intuitive criterion.

Proof of Lemma 3. To construct such a deviation for the high type, choose $\alpha < \alpha_L$ and set c such that

$$\begin{aligned}(1-\alpha)c &= (1-\alpha_L)c_L + \frac{1}{2}\left[(1-\alpha)m(\Delta^{-1}(\alpha\sqrt{\lambda_L})) - (1-\alpha_L)[m(\Delta^{-1}(\alpha_L\sqrt{\lambda_L}))\right] \\ &\quad + \frac{1}{2}\left[(1-\alpha)m(\Delta^{-1}(\alpha\sqrt{\lambda_H})) - (1-\alpha_L)[m(\Delta^{-1}(\alpha_L\sqrt{\lambda_H}))\right]. \end{aligned} \tag{34}$$

Recall that

$$\frac{1}{\lambda_L}\pi_L(\alpha, c) = (1-\alpha)m(\Delta^{-1}(\alpha\sqrt{\lambda_L})) - (1-\alpha)c.$$

Then substituting (34) into this and rearranging terms give

$$\begin{aligned}\frac{1}{\lambda_L}\pi_L(\alpha, c) &= (1-\alpha_L)m(\Delta^{-1}(\alpha_L\sqrt{\lambda_L})) - (1-\alpha_L)c_L \\ &\quad -\frac{1}{2}(1-\alpha_L)\left[m(\Delta^{-1}(\alpha_L\sqrt{\lambda_L})) - m(\Delta^{-1}(\alpha_L\sqrt{\lambda_H}))\right] \\ &\quad +\frac{1}{2}(1-\alpha)\left[m(\Delta^{-1}(\alpha\sqrt{\lambda_L})) - m(\Delta^{-1}(\alpha\sqrt{\lambda_H}))\right] \\ &< (1-\alpha_L)m(\Delta^{-1}(\alpha_L\sqrt{\lambda_L})) - (1-\alpha_L)c_L, \end{aligned} \tag{35}$$

where the last inequality follows from Assumption 2. Then, (35) can equivalently be written as

$$\pi_L(\alpha, c) < (1-\alpha_L)\lambda_L\left[m(\Delta^{-1}(\alpha_L\sqrt{\lambda_L})) - c_L\right] = \pi_L(\alpha_L, c_L).$$

Similarly we have that

$$\frac{1}{\lambda_H}\pi_H(\alpha, c) = (1-\alpha)m(\Delta^{-1}(\alpha\sqrt{\lambda_H})) - (1-\alpha)c.$$

Then substituting (34) into this and rearranging terms give

$$\begin{aligned}\frac{1}{\lambda_H}\pi_H(\alpha, c) &= (1-\alpha_L)m(\Delta^{-1}(\alpha_L\sqrt{\lambda_H})) - (1-\alpha_L)c_L \\ &\quad +\frac{1}{2}(1-\alpha_L)\left[m(\Delta^{-1}(\alpha_L\sqrt{\lambda_L})) - m(\Delta^{-1}(\alpha_L\sqrt{\lambda_H}))\right] \\ &\quad -\frac{1}{2}(1-\alpha)\left[m(\Delta^{-1}(\alpha\sqrt{\lambda_L})) - m(\Delta^{-1}(\alpha\sqrt{\lambda_H}))\right] \\ &> (1-\alpha_L)m(\Delta^{-1}(\alpha_L\sqrt{\lambda_H})) - (1-\alpha_L)c_L, \end{aligned} \tag{36}$$

where the inequality follows from Assumption 2. Clearly, (36) is equivalent to

$$\pi_H(\alpha, c) > (1-\alpha_L)\lambda_H\left[m(\Delta^{-1}(\alpha_L\sqrt{\lambda_H})) - c_L\right],$$

where the right hand side is equal to $\pi_H(\alpha_L, c_L)$ by definition, which, in turn, is equal to $\pi_H(\alpha_H, c_H)$ by assumption. Thus, we have

$$\pi_H(\alpha, c) > \pi_H(\alpha_H, c_H).$$

Moreover, since (α_L, c_L) and (α_H, c_H) are equilibrium outcomes, the service provider will assign positive probability to the originator being a low type upon seeing the contract offer (α_L, c_L), that is, $1 - \mu(\alpha_L, c_L) > 0$. Given the service provider accepts the contract offer (α_L, c_L) in equilibrium, we must have

$$\mu(c_L, \alpha_L)\left[(1-\alpha_L)c_L - k(1+\frac{\Delta^{-1}(\alpha_L\sqrt{\lambda_H})}{\sqrt{\lambda_H}})\right]+(1-\mu(c_L, \alpha_L))\left[(1-\alpha_L)c_L - k(1+\frac{\Delta^{-1}(\alpha_L\sqrt{\lambda_L}))}{\sqrt{\lambda_L}}\right] \geq 0.$$

Then we have that

$$(1-\alpha_L)c_L - k\left(1+\frac{\Delta^{-1}(\alpha_L\sqrt{\lambda_H})}{\sqrt{\lambda_H}}\right) > 0,$$

which follows since

$$\begin{aligned} 0 \;\leq\; & \mu(c_L, \alpha_L)\left[(1-\alpha_L)c_L - k(1+\frac{\Delta^{-1}(\alpha_L\sqrt{\lambda_H})}{\sqrt{\lambda_H}})\right] \\ & +(1-\mu(c_L, \alpha_L))\left[(1-\alpha_L)c_L - k(1+\frac{\Delta^{-1}(\alpha_L\sqrt{\lambda_L}))}{\sqrt{\lambda_L}}\right] \\ < \;& (1-\alpha_L)c_L - k\left(1+\frac{\Delta^{-1}(\alpha_L\sqrt{\lambda_H})}{\sqrt{\lambda_H}}\right). \end{aligned}$$

Then we also have that

$$(1-\alpha)c - k\left(1+\frac{\Delta^{-1}(\alpha\sqrt{\lambda_H})}{\sqrt{\lambda_H}}\right) \geq 0$$

for $\alpha < \alpha_L$ sufficiently close to α_L. Thus,

$$(1-\alpha)c\lambda_H - k\left(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha\sqrt{\lambda_H})\right) \geq 0.$$

The proof is then complete by Lemma 2. Finally, note that nowhere in this proof we assumed $(c_L, \alpha_L) \neq (c_H, \alpha_H)$.Thus, this result can be used to construct deviations from a pooling equilibrium as well. ■

Next, we prove Proposition 4.

Proof of Proposition 4. We follow the 4-step approach outlined next.

Step 1: Characterize a perfect Bayesian equilibrium as a solution to an optimization problem.

Step 2: Characterize/solve the optimization problem identified in Step 1.

Step 3: Show that the equilibrium characterized in Step 2 satisfies the intuitive criterion.

Step 4: Check that the only perfect Bayesian separating equilibrium satisfying the intuitive criterion is the one characterized in Step 2.

Step 1. Here we first focus on a specific separating equilibrium where the out of equilibrium beliefs are defined as follows: The service provider believes the originator to be a low type if he sees a contact offer different than (c_H, α_H), the equilibrium contract offer of the high type originator. We will eventually show that the equilibrium we derived under this belief system is the unique (pure strategy) perfect Bayesian equilibrium satisfying the intuitive criterion without any specific assumptions on the belief system. To this end, we first derive the constraints which have to be satisfied by a pair $\{(c_L, \alpha_L), (c_H, \alpha_H)\}$ of separating equilibrium contract offers. Consider the contract choice of the low type under the belief system μ that assigns probability zero to the originator being a high type upon seeing a contract offer other than (c_H, α_H). Then, the contract offer (c_L, α_L) is sequentially rational for the low type if it yields a profit that is better than any possible deviation. If the low type deviates to a contract $(c, \alpha) \neq (c_H, \alpha_H)$, then the service provider believes she is a low type. Thus, if (c, α) is accepted by the service provider, the low type gets profit $\pi_L(c, \alpha)$. Therefore, the best deviation for the low type to a contract $(c, \alpha) \neq (c_H, \alpha_H)$, then solves the following maximization problem: Choose the contract offer (c, α) so as to

$$\begin{aligned}
&\text{maximize} \quad \pi_L(c, \alpha) \\
&\text{subject to} \\
&(1-\alpha)c\lambda_L - k(\lambda_L + \sqrt{\lambda_L}\Delta^{-1}(\alpha\sqrt{\lambda_L})) \geq 0.
\end{aligned}$$

The constraint in this maximization problem always binds. Thus, the problem is the same as maximizing the surplus in the system. The solution is given by the first best level $\alpha = \alpha_L^*$ as in (5)-(6), and we have

$$(1-\alpha_L^*)c_L^* = k\left(1 + \frac{\Delta^{-1}(\alpha_L^*\sqrt{\lambda_L})}{\sqrt{\lambda_L}}\right).$$

Therefore, any deviation for the low type to a contract $(c, \alpha) \neq (c_H, \alpha_H)$ yields less profit for the low type. Then, for $(c_L, \alpha_L) = (c_L^*, \alpha_L^*)$ to be sequentially rational, it should also give higher profits for the low type than the equilibrium contract offer (c_H, α_H) of the high type. That is, we must have

$$\pi_L(c_L^*, \alpha_L^*) \geq (1-\alpha_H)\lambda_L[m(\Delta^{-1}(\alpha_H\sqrt{\lambda_L})) - c_H]. \tag{37}$$

We next consider the sequential rationality of the contract choice (c_H, α_H) of the high type. If the high type deviates from (c_H, α_H), under the belief system μ, the call center will believe that she is a low type. Thus, the best such deviation will solve the following problem: Choose the contract offer (c, α) so as to

$$\text{maximize} \;\; \pi_H(c, \alpha) \tag{38}$$

subject to

$$(1-\alpha)c\lambda_L - k(\lambda_L + \sqrt{\lambda_L}\Delta^{-1}(\alpha\sqrt{\lambda_L})) \geq 0. \tag{39}$$

Note that the constraint (39) reflects the fact that the service provider accepts the contract believing that it is offered by a low type originator. It is easy to see that the constraint (39) will always bind in optimum contract solving the problem (38)-(39). Thus, we can equivalently state (38)-(39) as follows: Choose the abandonment probability α so as to

$$\text{maximize } (1-\alpha)m(\Delta^{-1}(\alpha\sqrt{\lambda_H})) - k(1 + \frac{\Delta^{-1}(\alpha\sqrt{\lambda_L})}{\sqrt{\lambda_L}}). \tag{40}$$

which, in turn, can also be posed as a problem of choosing the standardized capacity β to

$$\text{maximize } (1 - \frac{\Delta(\beta)}{\sqrt{\lambda_H}})m(\beta) - k(1 + \frac{\widehat{\beta}}{\sqrt{\lambda_L}}), \tag{41}$$

where $\widehat{\beta}$ is defined as a function of β implicitly through the relation

$$\frac{\Delta(\beta)}{\sqrt{\lambda_H}} = \frac{\Delta(\widehat{\beta})}{\sqrt{\lambda_L}}. \tag{42}$$

The first order condition associated with the problem (41) can be written as

$$\left[m(\beta)(1 - \frac{\Delta(\beta)}{\sqrt{\lambda_H}})\right]' = \frac{k}{\sqrt{\lambda_L}}\frac{\partial\widehat{\beta}}{\partial\beta}, \tag{43}$$

where $\partial\widehat{\beta}/\partial\beta$ can be calculated using the implicit function theorem. More specifically, define the function φ such that

$$\varphi(\widehat{\beta}, \beta) = \Delta(\widehat{\beta}) - \sqrt{\frac{\lambda_L}{\lambda_H}}\Delta(\beta) = 0.$$

We have that

$$\partial\widehat{\beta}/\partial\beta = -\frac{\partial\varphi/\partial\beta}{\partial\varphi/\partial\widehat{\beta}} = \sqrt{\frac{\lambda_L}{\lambda_H}}\frac{\Delta'(\beta)}{\Delta'(\widehat{\beta})}.$$

Substituting this into (43), the first order condition becomes

$$\left[m(\beta)(1 - \frac{\Delta(\beta)}{\sqrt{\lambda_H}})\right]' = \frac{k}{\sqrt{\lambda_H}}\frac{\Delta'(\beta)}{\Delta'(\widehat{\beta})}, \tag{44}$$

where the right hand side is strictly less than $k/\sqrt{\lambda_H}$ since $\widehat{\beta} > \beta$ and $\Delta'(\widehat{\beta}) > \Delta'(\beta)$. Then comparing (44) with (5) and using Assumption 1, we see that the optimum solution $\widetilde{\beta}$ to the problem (42) satisfies $\widetilde{\beta} > \beta^*_H$, where β^*_H is the first best level. This implies that the optimal abandonment probability $\widetilde{\alpha}$ for the problem 38)-(39), or equivalently for the problem (40), satisfies

$$\widetilde{\alpha} = \frac{\Delta\left(\widetilde{\beta}\right)}{\sqrt{\lambda_H}} < \alpha^*_H, \tag{45}$$

where α^*_H is the first-best abandonment probability. Thus, the sequential rationality of (c_H, α_H) requires that

$$\pi_H(c_H, \alpha_H) \geq \lambda_H \, \widetilde{g}(\widetilde{\alpha}), \tag{46}$$

where

$$\widetilde{g}(\alpha) = (1-\alpha)m(\Delta^{-1}(\alpha\sqrt{\lambda_H})) - k(1 + \frac{\Delta^{-1}(\alpha\sqrt{\lambda_L})}{\sqrt{\lambda_L}}).$$

The last constraint the contract (c_H, α_H) has to satisfy is that it has to be accepted by the call center. Since $\mu(c_H, \alpha_H) = 1$, this translate into the constraint

$$(1-\alpha_H)c_H\lambda_H - k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha_H\sqrt{\lambda_H})) \geq 0. \tag{47}$$

In summary, for (c_L, α_L) and (c_H, α_H) to be equilibrium contract offers we must have $(c_L, \alpha_L) = (c^*_L, \alpha^*_L)$ and that (c_H, α_H) should solve the following problem: Choose (c, α) so as to

$$\text{maximize} \;\; \pi_H(c, \alpha) \tag{48}$$

subject to

$$(1-\alpha)c\lambda_H - k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha\sqrt{\lambda_H})) \geq 0, \tag{49}$$

$$\pi_H(c, \alpha) \geq \lambda_H\widetilde{g}(\widetilde{\alpha}), \tag{50}$$

$$\pi_L(c^*_L, \alpha^*_L) \geq (1-\alpha)\lambda_L[m(\Delta^{-1}(\alpha\sqrt{\lambda_L})) - c]. \tag{51}$$

The problem (48)-(51) characterizes the equilibrium contract offers, completing Step 1.

Step 2. We now focus on solving the problem (48)-(51). Clearly, we can reformulate the problem (48)-(51) as one of choosing $(1-\alpha_H)c_H$ and α_H. Without loss of generality, set

$$\varepsilon_H = k(1 + \frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_L})}{\sqrt{\lambda_L}}) - (1-\alpha_H)c_H. \tag{52}$$

Then,we can rewrite the problem as one of choosing ε_H and α_H optimally. Substituting (52) into (49)-(51), we can write the constraints as follows:

$$\varepsilon_H \leq h(\alpha_H),$$

$$\varepsilon_H \geq \widetilde{g}(\widetilde{\alpha}) - \widetilde{g}(\alpha_H),$$

$$\varepsilon_H \leq g_L(\alpha_L) - g_L(\alpha_H),$$

where $h(\cdot)$ and $g_L(\cdot)$ are defined as

$$h(\alpha) = k\left[\frac{\Delta^{-1}(\alpha\sqrt{\lambda_L})}{\sqrt{\lambda_L}} - \frac{\Delta^{-1}(\alpha\sqrt{\lambda_H})}{\sqrt{\lambda_H}}\right], \quad g_L(\alpha) = (1-\alpha)[m(\Delta^{-1}(\alpha\sqrt{\lambda_L}))] - k\left(1 + \frac{\Delta^{-1}(\alpha\sqrt{\lambda_L})}{\sqrt{\lambda_L}}\right).$$

Then, we can equivalently express the problem (48)-(51) as choosing ε_H and α_H so as to

$$\begin{aligned}
&\text{maximize } \widetilde{g}(\alpha_H) + \varepsilon_H \\
&\text{subject to} \\
&\varepsilon_H \leq h(\alpha_H), \\
&\widetilde{g}(\widetilde{\alpha}) - \widetilde{g}(\alpha_H) \leq \varepsilon_H \leq g_L(\alpha_L) - g_L(\alpha_H).
\end{aligned}$$

We further simplify the problem as follows: Choose ε_H and α_H to

$$\text{maximize } \widetilde{g}(\alpha_H) + \varepsilon_H \tag{53}$$

subject to

$$\varepsilon_H + \widetilde{g}(\alpha_H) \leq g_H(\alpha_H), \tag{54}$$

$$\widetilde{g}(\widetilde{\alpha}) \leq \varepsilon_H + \widetilde{g}(\alpha_H) \leq g_L(\alpha_L) - g_L(\alpha_H) + \widetilde{g}(\alpha_H), \tag{55}$$

where $g_H(\alpha) = h(\alpha) + \widetilde{g}(\alpha) = (1-\alpha)[m(\Delta^{-1}(\alpha\sqrt{\lambda_H}))] - k\left(1 + \frac{\Delta^{-1}(\alpha\sqrt{\lambda_H})}{\sqrt{\lambda_H}}\right)$. That is, we consider the following maximization problem in which we choose ζ, α_H so as to

$$\text{maximize } \zeta \tag{56}$$

subject to

$$\zeta \leq g_H(\alpha_H), \tag{57}$$

$$\widetilde{g}(\widetilde{\alpha}) \leq \zeta \leq f(\alpha_H), \tag{58}$$

where $f(\alpha) = g_L(\alpha_L) - g_L(\alpha) + \widetilde{g}(\alpha)$. To characterize the solution to the problem (56)-(58), we make the following observations.

Observation 1:

(i) $g_H(\cdot)$ achieves its unique maximum at α_H^*.

(ii) $g_H(\alpha) > \widetilde{g}(\alpha)$ for all $\alpha \in (0,1]$. Moreover, $g_H(\alpha) > g_H(\widetilde{\alpha}) > \widetilde{g}(\widetilde{\alpha})$ for $\widetilde{\alpha} < \alpha < \alpha_H^*$.

(iii) There exists a unique $\underline{\alpha} < \widetilde{\alpha}$ such that $g_H(\underline{\alpha}) = \widetilde{g}(\widetilde{\alpha})$.

It is easy to see that part (i) follows from Assumption 1 and the first order condition (**??**). We have $g_H(\alpha) > \widetilde{g}(\alpha)$ for all $\alpha \in (0,1]$ since Δ is decreasing and $\lambda_H > \lambda_L$. Then, $g_H(\alpha) > g_H(\widetilde{\alpha}) >$

$\widetilde{g}(\widetilde{\alpha})$ for $\widetilde{\alpha} < \alpha < \alpha_H^*$ from Part (*i*). Part (*iii*) simply follows from Part (*ii*) and monotonicity of g_H.

Observation 2: $f(\cdot)$ is decreasing with $f(\alpha_H^*) < g_H(\alpha_H^*)$ and $f(\underline{\alpha}) > g_H(\underline{\alpha}) = \widetilde{g}(\widetilde{\alpha})$.

Assumption 2 implies that $f(\cdot)$ is decreasing. We have $f(\alpha_H^*) = g_L(\alpha_L) - g_L(\alpha_H^*) + \widetilde{g}(\alpha_H^*) < g_H(\alpha_H^*)$ since $g_L(\alpha_L) < g_H(\alpha_H^*)$ and $g_L(\alpha_H^*) > \widetilde{g}(\alpha_H^*)$. Finally, $f(\underline{\alpha}) > g_H(\underline{\alpha}) = \widetilde{g}(\widetilde{\alpha})$ since $f(\widetilde{\alpha}) > \widetilde{g}(\widetilde{\alpha})$, $f(\cdot)$ if decreasing and $\underline{\alpha} < \widetilde{\alpha}$.

Observation 3: There exists a unique $\alpha_H \in (\underline{\alpha}, \alpha_H^*)$ such that $f(\alpha_H) = g_H(\alpha_H)$, which also satisfies $f(\alpha_H) > \widetilde{g}(\widetilde{\alpha})$. Moreover, (c_H, α_H), where

$$c_H = \frac{k}{(1-\alpha_H)}\left(1 + \frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}}\right), \tag{59}$$

is the unique optimal solution of (48)-(51).

To see why α_H and c_H as in (59) is the unique optimal solution of (48)-(51), recall that the constraint (57), or equivalently (54), binds at an optimum solution. Then, since the constraint (57) was obtained from (50) through the relation (52), ε_H denote the slack in (50). The fact that (57) binds at the optimum implies that $\varepsilon_H = 0$ and hence, α_H and c_H as in (59) constitute an equilibrium contract offer for the high type.

Therefore, (c_L^*, α_L^*) and (c_H, α_H) as characterized in Observation 3 are the equilibrium contract offers by the low and high types, under the belief system μ. The next step is to this equilibrium satisfies the intuitive criterion.

Step 3.

Let (c_L^*, α_L^*) and (c_H, α_H) denote the contract offers by the high and low type originators in the separating equilibrium of Step 2. The high and low type originators earn the following profits:

$$\begin{aligned}
\pi_H(c_H, \alpha_H) &= (1-\alpha_H)[m(\Delta^{-1}(\alpha_H\sqrt{\lambda_H})) - c_H], \\
\pi_L(c_L^*, \alpha_L^*) &= (1-\alpha_L^*)m(\Delta^{-1}(\alpha_L^*\sqrt{\lambda_L})) - k(\lambda_L + \sqrt{\lambda_L}\Delta^{-1}(\alpha_L^*\sqrt{\lambda_L})).
\end{aligned}$$

We will show that there for both the low and high types, there exists no deviation that would violate the intuitive criterion. We first consider the potential deviations by the low type. Since we have

$$\pi_L(c_L^*, \alpha_L^*)) > 0,$$

the low type would not deviate to a contract that would be rejected by the call center. Put more formally, for there to be a deviation (c, α) for the low type such that

$$\mathrm{Min}_{s\in S^*(\Theta^*(c,\alpha),\ (c,\alpha))}\ u_L(c,\alpha;s) > \pi_L(c_L^*, \alpha_L^*)),$$

we should have that $S^*(\Theta^*(c,\alpha),\ (c,\alpha)) = \{\text{Accept}\}$. Suppose such a deviation (c,α) exists. This immediately implies that $u_L(c,\alpha;\text{Accept}) > u_L(c_L^*,\alpha_L^*;\text{Accept})$. Thus, $\{L\} \subset \Theta^*(c,\alpha)$ if $\Theta^*(c,\alpha) \neq \varnothing$. Recall that by definition, the set $S^*(\Theta^*(c,\alpha),\ (c,\alpha))$ contains the equilibrium responses by the call center to the contract choice (c,α) for some beliefs that assign positive probability to types only in the set $\Theta^*(c,\alpha)$. Then, to have $S^*(\Theta^*(c,\alpha),(c,\alpha)) = \{\text{Accept}\}$, the contract (c,α) has to satisfy

$$\lambda_L(1-\alpha)c - k(\lambda_L + \sqrt{\lambda_L}\Delta^{-1}(\alpha\sqrt{\lambda_L})) \geq 0,$$

in which case we have

$$u_L(c,\alpha;\text{Accept}) \leq u_L(c_L,\alpha_L;\text{Accept}),$$

as α_L^* is the first best level of abandonment rate for the low type, contradicting the assumption that (c,α) is a deviation that yields higher for the low type originator.

Second we consider potential deviations by the high type originator. Suppose there exists a contract offer $(c,\alpha) \neq (c_H,\alpha_H)$ by the high type such that

$$\text{Min}_{s\in S^*(\Theta^*(c,\alpha),\ (c,\alpha))}\ u_H(c,\alpha;s) > u_H(c_H,\alpha_H).$$

Since (c_H,α_H) gives the high type originator nonnegative profits, we should have $S^*(\Theta^*(c,\alpha),$ $(c,\alpha)) = \{\text{Accept}\}$. Then, since $u_H(c,\alpha;\text{Accept}) > u_H(c_H,\alpha_H;\text{Accept})$, we have $\{H\} \subset \Theta^*(c,\alpha)$. Now consider two cases that might potentially arise. In the first case, we have $\Theta^*(c,\alpha) = \{H\}$. If $\Theta^*(c,\alpha) = \{H\}$, to have $S^*(\Theta^*(c,\alpha),\ (c,\alpha)) = \{\text{Accept}\}$, it should be that

$$\lambda_H(1-\alpha)c - k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha\sqrt{\lambda_H})) \geq 0.$$

Moreover, by the definition of the set $\Theta^*(c,\alpha) = \{H\}$, we must also have

$$\begin{aligned}(1-\alpha_L^*)[m(\Delta^{-1}(\alpha_L^*\sqrt{\lambda_L})) - c_L^*] \quad &> \quad (1-\alpha)[m(\Delta^{-1}(\alpha\sqrt{\lambda_L})) - c],\\ (1-\alpha_H)[m(\Delta^{-1}(\alpha\sqrt{\lambda_H})) - c_H] \quad &< \quad (1-\alpha)[m(\Delta^{-1}(\alpha\sqrt{\lambda_H})) - c],\end{aligned}$$

However, this would contradict the sequential rationality of the contract choice (c_H,α_H) in the separating equilibrium Step 2. Thus, there exits no deviation (c,α) for the high type such that $\Theta^*(c,\alpha) = \{H\}$ and the intuitive criterion is violated.

In the second case, we have $\Theta^*(c,\alpha) = \{H,L\}$. If $\Theta^*(c,\alpha) = \{H,L\}$, then to have $S^*(\Theta^*(c,\alpha),$ $(c,\alpha)) = \{\text{Accept}\}$, we must have

$$\lambda_L(1-\alpha)c - k(\lambda_L + \sqrt{\lambda_L}\Delta^{-1}(\alpha\sqrt{\lambda_L})) \geq 0. \tag{60}$$

Then, the best possible profit a the high type originator could get from a deviation (c, α) that satisfy (60) is

$$(1-\widetilde{\alpha})\lambda_H \left[m(\Delta^{-1}(\widetilde{\alpha}\sqrt{\lambda_H})) - k(1 + \frac{\Delta^{-1}(\widetilde{\alpha}\sqrt{\lambda_L})}{\sqrt{\lambda_L}} \right],$$

where $\widetilde{\alpha}$ is given by (45). From (46), the deviation (c, α) yields less profits for the high type originator than $u_H(c_H, \alpha_H)$. Thus, no deviation exists for the high type that would violate the intuitive criterion and the equilibrium in Step 2 satisfies the intuitive criterion.

Step 4. We now show that the separating equilibrium characterized in Step 2 is the unique outcome that might arise as a separating equilibrium satisfying the intuitive criterion.

To check this, let $\{(c_H, \alpha_H), (c_L, \alpha_L)\}$ be an arbitrary separating equilibrium contract offers satisfying the intuitive criterion. First, we show that we must have

$$(1-\alpha_L)c_L = k\left(1 + \frac{\Delta^{-1}(\alpha_L\sqrt{\lambda_L})}{\sqrt{\lambda_L}}\right). \tag{61}$$

That is, the contract offer of the low type originator does not leave any surplus to the service provider. Second, we show that we must have $\alpha_L = \alpha_L^*$. Finally, we show that α_H and c_H are as characterized in Step 2. In particular, they satisfy

$$\alpha_H < \alpha_H^* \text{ and } (1-\alpha_H)c_H = k\left(1 + \frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}}\right).$$

To show that (61) holds, we argue by contradiction. Note that in a separating equilibrium, upon seeing the low type's contract offer (c_L, α_L), the service provider knows that she is low type. Since the service provider accepts the contract offered in equilibrium, we must have that

$$(1-\alpha_L)c_L \geq k\left(1 + \frac{\Delta^{-1}(\alpha_L\sqrt{\lambda_L})}{\sqrt{\lambda_L}}\right).$$

Thus, we only need to rule out the possibility that

$$(1-\alpha_L)c_L > k\left(1 + \frac{\Delta^{-1}(\alpha_L\sqrt{\lambda_L})}{\sqrt{\lambda_L}}\right). \tag{62}$$

Suppose (62) holds. Then, $(c_L - \epsilon, \alpha_L)$ for $\epsilon > 0$ sufficiently small is a deviation that violate the intuitive criterion. To see this, it suffices to check that $\pi_H(\alpha_H, c_H) > \pi_H(\alpha_L, c_L - \epsilon)$, cf. Lemma 2. By incentive compatibility, we must have $\pi_H(\alpha_H, c_H) \geq \pi_H(\alpha_L, c_L)$. Then, if $\pi_H(\alpha_H, c_H) > \pi_H(\alpha_L, c_L)$, we also have $\pi_H(\alpha_H, c_H) > \pi_H(\alpha_L, c_L - \epsilon)$ for ϵ small enough. Therefore, we restrict attention to the case $\pi_H(\alpha_H, c_H) = \pi_H(\alpha_L, c_L)$, that is, the high type is indifferent between his contract and that of the low type, but then one can invoke Lemma 3 to find a deviation (c, α) for the high type originator such that

$$\pi_H(\alpha, c) > \pi_H(\alpha_H, c_H),\ \pi_L(\alpha_L, c_L) > \pi_L(\alpha, c) \text{ and } (1-\alpha)c\lambda_H - k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha\sqrt{\lambda_H})) \geq 0,$$

which is a contradiction. Thus, the low type extracts all the surplus from the service provider in any separating equilibrium satisfying the intuitive criterion.

Next, we prove that $\alpha_L = \alpha_L^*$, for which we use the fact that

$$(1-\alpha_L)c_L = k\left(1 + \frac{\Delta^{-1}(\alpha_L\sqrt{\lambda_L})}{\sqrt{\lambda_L}}\right). \tag{63}$$

We argue by contradiction, i.e. suppose $\alpha_L \neq \alpha_L^*$. If $\pi_H(\alpha_H, c_H) > \pi_H(\alpha_L, c_L)$, then moving slightly towards α_L^* would be a deviation for the low type that violates the intuitive criterion. On the other hand, if $\pi_H(\alpha_H, c_H) = \pi_H(\alpha_L, c_L)$, the argument is the same as the one immediately above. Namely, the high type can find a deviation that violates the intuitive criterion. Thus, we have $\alpha_L = \alpha_L^*$.

Finally, we prove that α_H and c_H are as characterized in Step 2. in particular, we shoe that they satisfy

$$\alpha_H < \alpha_H^* \quad \text{and} \quad (1-\alpha_H)c_H = k\left(1 + \frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}}\right).$$

To this end, let $(\overline{\alpha}_H, \overline{c}_H)$ denote the specific contract offer of the high type characterized in Step 2. Since we already know that $\alpha_L = \alpha_L^*$ and c_L is as in (63), we need to only consider deviations by the high type. We argue by contradiction. Suppose $(\alpha_H, c_H) \neq (\overline{\alpha}_H, \overline{c}_H)$. Then, it follows from the proof of Step 2 that if

$$\alpha_H > \overline{\alpha}_H \quad \text{and} \quad (1-\alpha_H)c_H = k(1 + \frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}}),$$

then the low type will prefer (α_H, c_H) over (α_L, c_L). Then it is easy to show that there exists a deviation for the low type violating the intuitive criterion as in Lemma 2. on the other hand if

$$\alpha_H > \overline{\alpha}_H \quad \text{and} \quad (1-\alpha_H)c_H > k(1 + \frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}}),$$

then the contract $(\alpha_H, c_H - \epsilon)$ for ϵ small enough yields strictly better profits for the high type, and we will have

$$(1-\alpha_H)\left(c_H - \epsilon\right) > k(1 + \frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}}).$$

Moreover, we would have $\pi_L\left(\alpha_H, c_H - \epsilon\right) < \pi_L\left(\alpha_L, c_L\right)$ if $\pi_L\left(\alpha_H, c_H\right) < \pi_L\left(\alpha_L, c_L\right)$ already. Therefore we restrict attention to the case $\pi_L\left(\alpha_H, c_H\right) = \pi_L\left(\alpha_L, c_L\right)$. Then, there exists a deviation (α, c) for the high type such that $\pi_H(\alpha_H, c_H) < \pi_H(\alpha, c)$, $\pi_L(\alpha_H, c_H) > \pi_L(\alpha, c)$ and $(1-\alpha)c\lambda_H - k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha\sqrt{\lambda_H})) \geq 0$, which follows from Lemma 3, and that violates the intuitive criterion by Lemma 2. Thus, we have proved that we cannot have $\alpha_H > \overline{\alpha}_H$.

We next show that we cannot have $\alpha_H < \overline{\alpha}_H$. If $\alpha_H < \overline{\alpha}_H$ and

$$(1-\alpha_H)c_H = k(1+\frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}}),$$

then the high type can increase α_H slightly and set

$$(1-\alpha_H)c_H = k(1+\frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}}),$$

which will yield strictly higher profits for her, while the low type would strictly prefer

$$\alpha_L = \alpha_L^* \text{ and } (1-\alpha_L^*)c_L = k(1+\frac{\Delta^{-1}(\alpha_L^*\sqrt{\lambda_L})}{\sqrt{\lambda_L}})$$

over the deviation of the high type since $\pi_L(\alpha_H, c_H) < \pi_L(\alpha_L, c_L)$ for $\alpha_H < \overline{\alpha}_H$ and $(1-\alpha_H)c_H = k(1+\frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}})$. Then what remains to show is that $\alpha_H < \overline{\alpha}_H$ and $(1-\alpha_H)c_H > k(1+\frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}})$ cannot be an equilibrium outcome in a separating equilibrium satisfying the intuitive criterion. If $\pi_L(\alpha_H, c_H) < \pi_L(\alpha_L, c_L)$, then $(\alpha_H, c_H - \epsilon)$ for ϵ small enough is a deviation for the high type that violates the intuitive criterion. Finally, observe that we cannot have $\pi_L(\alpha_H, c_H) = \pi_L(\alpha_L, c_L)$ if $\alpha_H < \overline{\alpha}_H$ and

$$(1-\alpha_H)c_H > k(1+\frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}})$$

since $\pi_L(\overline{\alpha}_H, \overline{c}_H) = \pi_L(\alpha_L, c_L)$. Thus, we have that $\alpha_H = \overline{\alpha}_H$.

Finally we show that

$$(1-\alpha_H)c_H = k(1+\frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}}).$$

We argue by contradiction again. Suppose that

$$(1-\alpha_H)c_H > k(1+\frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}}).$$

Then, $\pi_L(\alpha_H, c_H) < \pi_L(\alpha_L, c_L)$ by the proof of Step 2. Hence, $(\alpha_H, c_H - \epsilon)$ for ϵ small enough is a deviation for the high type that violates the intuitive criterion. Thus, we have $\alpha_H = \overline{\alpha}_H$ and

$$(1-\alpha_H)c_H = k(1+\frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}}),$$

which prove that α_H and c_H are as characterized in Step 2. ■

Proof of Proposition 5. We argue by contradiction. Suppose that there exists a pooling equilibrium that satisfies the intuitive criterion. Let $(\widetilde{c}, \widetilde{\alpha})$ denote an arbitrary pooling equilibrium contract offer with the belief system μ. From Bayes' rule, we have $\mu(\widetilde{c}, \widetilde{\alpha}) = p$. Note that $(\widetilde{c}, \widetilde{\alpha})$ has to be accepted by the service provider since otherwise some type would find it profitable to deviate and offer a contract that is accepted. Since $(\widetilde{c}, \widetilde{\alpha})$ is accepted in the equilibrium by the service provider, we have

$$p\left[(1-\widetilde{\alpha})\widetilde{c}\lambda_H - k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\widetilde{\alpha}\sqrt{\lambda_H}))\right] + (1-p)\left[(1-\widetilde{\alpha})\widetilde{c}\lambda_L - k(\lambda_L + \sqrt{\lambda_L}\Delta^{-1}(\widetilde{\alpha}\sqrt{\lambda_L}))\right] \geq 0.$$

Next, we show that the high type has a deviation that violates the intuitive criterion. It is easy to see that there exists a deviation (c, α) for the high type as in Lemma 3 deviation contract $(\widetilde{c}, \widetilde{\alpha})$ such that

$$\pi_H(c,\alpha) > \pi_H(\widetilde{c},\widetilde{\alpha}),\ \pi_L(\widetilde{c},\widetilde{\alpha}) > \pi_L(c,\alpha) \text{ and } (1-\alpha)c\lambda_H - k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha\sqrt{\lambda_H})) \geq 0,$$

which is a contradiction. Thus, there exists no pooling equilibrium that satisfies the intuitive criterion. ■

Proof of Proposition 3. The "if part" of Proposition 3 was established in Lemma 2. We now prove the "only if" part of Proposition 3. Namely, we prove that if for a perfect Bayesian equilibrium with contract offers (α_L, c_L) and (α_H, c_H), there exists no type $i \in \{L, H\}$ and no deviation contract (α, c) such that

$$\pi_i(\alpha,c) > \pi_i(\alpha_i,c_i),\ \pi_j(\alpha_j,c_j) > \pi_j(\alpha,c) \text{ and } (1-\alpha)c\lambda_i - k(\lambda_i + \sqrt{\lambda_i}\Delta^{-1}(\alpha\sqrt{\lambda_i})) \geq 0 \text{ for } j \neq i,$$

then the perfect Bayesian equilibrium with contract offers (α_L, c_L) and (α_H, c_H) does not violate the intuitive criterion. We have shown in Step 4 of the proof of Proposition 4 that the only perfect Bayesian separating equilibrium for which there exists no deviations as in Lemma 2 is the one characterized in Proposition 4. Moreover, we have also shown that the equilibrium characterized in Proposition 4 satisfies the intuitive criterion, This establishes that if for a perfect Bayesian equilibrium, there exits no deviations as in Lemma 2, then the perfect Bayesian equilibrium satisfies the intuitive criterion. ■

C Proofs in Section 5

First note , that in the screening scenario, the service provider's problem is: Choose $\{(\alpha_L, c_L, T_L), (\alpha_H, c_H, T_H)\}$ so as to

$$\text{maximize } p\,[T_H + c_H\lambda_H(1-\alpha_H) - k(\lambda_H + \sqrt{\lambda_H}\beta(\alpha_H,\lambda_H))] + (1-p)\,[T_L + c_L\lambda_L(1-\alpha_L) - k(\lambda_L + \sqrt{\lambda_L}\beta(\alpha_L,\lambda_L))]$$

subject to

$$\lambda_H(1-\alpha_H)[m(\beta(\alpha_H,\lambda_H)) - c_H] - T_H \geq 0, \quad (\text{IR}_H)$$

$$\lambda_L(1-\alpha_L)[m(\beta(\alpha_L,\lambda_L)) - c_L] - T_L \geq 0, \quad (\text{IR}_L)$$

$$\lambda_H(1-\alpha_H)[m(\beta(\alpha_H,\lambda_H)) - c_H] - T_H \geq \lambda_H(1-\alpha_L)[m(\beta(\alpha_L,\lambda_H)) - c_L] - T_L, \quad (\text{IC}_H)$$

$$\lambda_L(1-\alpha_L)[m(\beta(\alpha_L,\lambda_L)) - c_L] - T_L \geq \lambda_L(1-\alpha_H)[m(\beta(\alpha_H,\lambda_L)) - c_H] - T_H, \quad (\text{IC}_L)$$

where the first two constraints enforce individual rationality while the last two are incentive compatibility constraints.

Proof of Proposition 6. We first consider the case when the following condition holds:

$$\lambda_H m(\beta(\alpha_H^*,\lambda_H)) \geq \lambda_L m(\beta(\alpha_H^*,\lambda_L)). \quad (64)$$

We will first analyze the screening problem of the service provider without the constraints. This will give us unique levels of the abandonment probabilities to be offered to the high and low type originators. Then, we will show that there exists (c_L, T_L) and (c_H, T_H) such that we can actually implement the optimal abandonment probabilities found previously, i.e. the incentive compatibility and individual rationality constraints are satisfied. This clearly gives us an optimal solution to the screening problem of the service provider. Finally, we characterize the possible values of (c_L, T_L) and (c_H, T_H) that can implement the optimal abandonment probabilities.

If we consider the screening problem of the service provider where the service provider extracts all the surplus from the high and low type originators and the incentive compatibility and individual rationality constraints are ignored, it is easy to see that the optimal abandonment probabilities are given uniquely by the first best levels, cf. Proposition 1. That is, the optimal abandonment probabilities are given as follows

$$\alpha_L = \alpha_L^* \text{ and } \alpha_H = \alpha_H^*. \quad (65)$$

Next, we prove that there exists (c_L, T_L) and (c_H, T_H) that satisfy the incentive compatibility and individual rationality constraints given (8). We propose the following values for (c_L, T_L) and

(c_H, T_H) and show that together with (65), the incentive compatibility and individual rationality constraints are satisfied. Let the high type originator be offered only a fixed fee, i.e. $c_H = 0$ and $T_H = \lambda_H(1-\alpha_H^*)m(\beta(\alpha_H^*, \lambda_H))$ and the low type originator be offered only per-call fee, i.e. $T_L = 0$ and $c_L = m(\beta(\alpha_L^*, \lambda_L))$. Then, (IC$_H$) is satisfied. IC$_L$ is also satisfied since

$$
\begin{aligned}
0 &\geq \lambda_L(1-\alpha_H)[m(\beta(\alpha_H^*, \lambda_L))] - T_H, \\
&= \lambda_L(1-\alpha_H)[m(\beta(\alpha_H^*, \lambda_L))] - \lambda_H(1-\alpha_H)[m(\beta(\alpha_H^*, \lambda_H))],
\end{aligned}
$$

since (64) holds. Thus, if (64) holds, we can implement the first-best allocation by the payment scheme: $T_L = 0$, $c_L = m(\beta(\alpha_L^*, \lambda_L))$ and $c_H = 0$, $T_H = \lambda_H(1-\alpha_H^*)m(\beta(\alpha_H^*, \lambda_H))$. Then, it is straightforward to see that (9)-(12) characterize the possible selection of (c_L, T_L) and (c_H, T_H) that would be optimal for the screening problem of the service provider together with $\alpha_L = \alpha_L^*$ and $\alpha_H = \alpha_H^*$.

Second, we consider what happens if (64) does not hold. A result similar to Lemma 1 proves that (IC$_L$) and (IR$_H$) binds. Then, we have

$$
T_H + \lambda_H (1-\alpha_H) c_H = \lambda_H (1-\alpha_H) m (\beta (\alpha_H, \lambda_H)), \tag{66}
$$

by (IR$_H$). Substituting (66) into (IC$_L$) and rearranging terms yields

$$
\begin{aligned}
T_L + \lambda_L (1-\alpha_L) c_L &= \lambda_L (1-\alpha_L) m (\beta (\alpha_L, \lambda_L)) \\
&\quad + [\lambda_H (1-\alpha_H) [m (\beta (\alpha_H, \lambda_H)) - c_H] - \lambda_L (1-\alpha_H) [m (\beta (\alpha_H, \lambda_L)) - c_H]]
\end{aligned}
\tag{67}
$$

Next we plug the expressions for $T_H + \lambda_H (1-\alpha_H) c_H$ and $T_L + \lambda_L (1-\alpha_L) c_L$ into the objective function of the maximization problem of the call center and solve for optimum abandonment probabilities. That is, we choose α_L and α_H so as to

$$
\begin{aligned}
&\text{maximize } p[\lambda_H(1-\alpha_H)m(\beta(\alpha_H, \lambda_H)) - k(\lambda_H + \sqrt{\lambda_H}\beta(\alpha_H, \lambda_H))] \\
&+(1-p)[\lambda_L(1-\alpha_L)m(\beta(\alpha_L, \lambda_L)) - k(\lambda_L + \sqrt{\lambda_L}\beta(\alpha_L, \lambda_L)) \\
&+[\lambda_H (1-\alpha_H) [m (\beta (\alpha_H, \lambda_H)) - c_H] - \lambda_L (1-\alpha_H) [m (\beta (\alpha_H, \lambda_L)) - c_H]]]
\end{aligned}
$$

It is easy to see that the optimizing level of α_L is equal to α_L^*, the first best level. To find the optimum α_H and c_H, we solve the following problem if we ignore the terms that do not depend on α_H and c_H: Choose α_H, c_H so as to

$$
\begin{aligned}
&\text{maximize } p[m(\beta(\alpha_H, \lambda_H))\lambda_H(1-\alpha_H) - k(\lambda_H + \sqrt{\lambda_H}\beta(\alpha_H, \lambda_H))] \\
&+(1-p)(1-\alpha_H)[\lambda_H m (\beta (\alpha_H, \lambda_H)) - \lambda_L m (\beta (\alpha_H, \lambda_L)) + (\lambda_L - \lambda_H) c_H].
\end{aligned}
$$

Since $\lambda_L - \lambda_H < 0$, the in the optimal solution we have $c_H = 0$. Then, the optimization with respect to α_H can equivalently be written as a problem of maximization with respect to β_H.

$$\begin{aligned}&\text{Maximize}_{\beta_H}\ p[m(\beta_H)\lambda_H(1-\frac{\Delta(\beta_H)}{\sqrt{\lambda_H}}) - k(\lambda_H + \sqrt{\lambda_H}\beta_H)]\\&+(1-p)(1-\frac{\Delta(\beta_H)}{\sqrt{\lambda_H}})\left[\lambda_H m(\beta_H) - \lambda_L m(\widehat{\beta}_H)\right],\end{aligned}$$

where $\widehat{\beta}_H$ is defined as a function of β_H through the relation

$$\frac{\Delta(\beta_H)}{\sqrt{\lambda_H}} = \frac{\Delta(\widehat{\beta}_H)}{\sqrt{\lambda_L}}.$$

The first order conditions give

$$\left[m(\beta_H)(1-\frac{\Delta(\beta_H)}{\sqrt{\lambda_H}})\right]' = \frac{k}{\sqrt{\lambda_H}} - \frac{(1-p)}{p}\left[(1-\frac{\Delta(\beta_H)}{\sqrt{\lambda_H}})[m(\beta_H) - \frac{\lambda_L}{\lambda_H}m(\widehat{\beta}_H)]\right]'.$$

By rearranging terms we get

$$\left[m(\beta_H)(1-\frac{\Delta(\beta_H)}{\sqrt{\lambda_H}})\right]'\left(1+\frac{(1-p)}{p}\frac{\lambda_H-\lambda_L}{\lambda_H}\right) = \frac{k}{\sqrt{\lambda_H}} - \frac{(1-p)}{p}\frac{\lambda_L}{\lambda_H}\left[(1-\frac{\Delta(\beta_H)}{\sqrt{\lambda_H}})[m(\beta_H) - m(\widehat{\beta}_H)]\right]'$$

where the second term on the right hand side $\left[(1-\frac{\Delta(\beta_H)}{\sqrt{\lambda_H}})[m(\beta_H) - m(\widehat{\beta}_H)]\right]' \geq 0$ by Assumption 2 as before. Then, by Assumption 1, we have $\beta_H \geq \beta_H^*$.

What remains is to check whether (IC_H) is satisfied in this solution. For that, we rewrite (IC_H) as follows since (IR_H) binds:

$$\begin{aligned}0 &\geq \lambda_H(1-\alpha_L)[m(\beta(\alpha_L,\lambda_H)) - c_L] - T_L,\\&= \lambda_H(1-\alpha_L)m(\beta(\alpha_L,\lambda_H)) - [T_L + \lambda_L(1-\alpha_L)c_L] - (\lambda_H-\lambda_L)(1-\alpha_L)c_L,\end{aligned}$$

where $T_L + \lambda_L(1-\alpha_L)c_L$ is given by (67). Then, (IC_H) is most likely to be satisfied when c_L is as large as possible. Thus, let $T_L = 0$, and (IC_H) becomes

$$\begin{aligned}0 &\geq (1-\alpha_L)m(\beta(\alpha_L,\lambda_H)) - (1-\alpha_L)c_L,\\&= (1-\alpha_L)m(\beta(\alpha_L,\lambda_H)) - (1-\alpha_L)\,m\,(\beta\,(\alpha_L,\lambda_L)) - \frac{\lambda_H}{\lambda_L}(1-\alpha_H)\,m\,(\beta\,(\alpha_H,\lambda_H))\\&\quad + (1-\alpha_H)\,m\,(\beta\,(\alpha_H,\lambda_L)),\end{aligned}$$

which is satisfied by Assumption 2 and the fact that $\lambda_H > \lambda_L$. ■

Proof of Proposition 7. We first consider the case when (15) holds. Then, it is easy to see that the contract offer characterized in part (i) of Proposition 7 is a separating equilibrium with a belief system that assigns probability zero for the originator being a high type upon seeing a contract offer other than (α_H, c_H, T_H). The proof that this equilibrium satisfies the intuitive criterion is very similar to the proof of Step 3 of Proposition 4. Then, we prove that this equilibrium is the unique separating equilibrium satisfying the intuitive criterion as follows: The argument that $\lambda_L(1-\alpha_L)c_L + T_L = k(\lambda_L + \sqrt{\lambda_L}\Delta^{-1}(\alpha_L\sqrt{\lambda_L}))$ and $\alpha_L = \alpha_L^*$ is very similar to the argument in Step 4 of the proof of Proposition 4, and is actually proved while we consider Part (ii). Then, we have $\lambda_H(1-\alpha_H)c_H + T_H = k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha_H\sqrt{\lambda_H}))$ and $\alpha_H = \alpha_H^*$ since otherwise there always exists a deviation for the high type that violates the intuitive criterion, either by moving towards α_H^* in terms of the abandonment probability or by decreasing the total payment to the service provider.

Next we consider the case when (15) does not hold. To that end, let $\{(c_H, \alpha_H, T_H), (c_L, \alpha_L, T_L)\}$ be an arbitrary separating equilibrium satisfying the intuitive criterion. We divide the proof into four steps. First, we show that $\lambda_L(1-\alpha_L)c_L + T_L = k(\lambda_L + \sqrt{\lambda_L}\Delta^{-1}(\alpha_L\sqrt{\lambda_L}))$. Second, we show that we must have $\alpha_L = \alpha_L^*$. Third, we prove that $c_H = 0$. Finally, we show that $\alpha_H < \alpha_H^*$ and $T_H = k(1 + \frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}})$ as in Proposition 7.

Suppose that $(1-\alpha_L)c_L + \frac{T_L}{\lambda_L} > k(1 + \frac{\Delta^{-1}(\alpha_L\sqrt{\lambda_L})}{\sqrt{\lambda_L}})$. Then, $(c_L, \alpha_L, T_L - \epsilon)$ for $\epsilon > 0$ small enough, is a deviation that violate the intuitive criterion. To see this, we need to check only that $\pi_H(\alpha_H, c_H, T_H) > \pi_H(\alpha_L, c_L, T_L - \epsilon)$. By incentive compatibility, we have $\pi_H(\alpha_H, c_H, T_H) \geq \pi_H(\alpha_L, c_L, T_L)$. Then, if $\pi_H(\alpha_H, c_H, T_H) > \pi_H(\alpha_L, c_L, T_L)$, for small enough ϵ, we still have $\pi_H(\alpha_H, c_H, T_H) > \pi_H(\alpha_L, c_L, T_L - \epsilon)$. Consider the other case, i.e. $\pi_H(\alpha_H, c_H, T_H) = \pi_H(\alpha_L, c_L, T_L)$. By a similar result to Lemma 3, we can show that there exists a deviation (α, c, T) for the high type such that $\pi_H(\alpha_H, c_H, T_H) < \pi_H(\alpha, c, T)$, $\pi_L(\alpha_L, c_L, T_L) > \pi_L(\alpha, c, T)$ and $(1-\alpha)c\lambda_H - k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha\sqrt{\lambda_H})) \geq 0$. This completes the proof first part, namely, the low type originator extracts all surplus from the call center in any separating equilibrium satisfying the intuitive criterion.

Second, we prove that $\alpha_L = \alpha_L^*$. Suppose not. Recall that we have $(1-\alpha_L)c_L = k(1 + \frac{\Delta^{-1}(\alpha_L\sqrt{\lambda_L})}{\sqrt{\lambda_L}})$ from the first part of the proof. If $\pi_H(\alpha_H, c_H, T_H) > \pi_H(\alpha_L, c_L, T_L)$, then moving slightly towards α_L^* would be a deviation for the low type that violates the intuitive criterion. If $\pi_H(\alpha_H, c_H, T_H) = \pi_H(\alpha_L, c_L, T_L)$, the argument is the same as the previous case. Thus, we have $\alpha_L = \alpha_L^*$.

Next, we prove that we should have $c_H = 0$. Suppose not. Then, we will construct a deviation $(c_H^{'}, \alpha_H^{'}, T_H^{'})$ that violates the intuitive criterion. If $(1-\alpha_H)c_H + \frac{T_H}{\lambda_H} > k(1 + \frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}})$, then let $c_H^{'} = 0$, $T_H^{'} = (1-\alpha_H)c_H\lambda_H + T_H - \epsilon$ and the low type would strictly prefer to stay at her optimal contract whereas the high type strictly prefers the deviation to the existing contract offer of her. Now consider the case $(1-\alpha_H)c_H + \frac{T_H}{\lambda_H} = k(1+\frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}})$. Then, the best level of α the high type could achieve is α_H^* but at α_H^* the low type strictly prefers to deviate. Then, let $c_H^{'} = 0$, $T_H^{'} = k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha_H^{'}\sqrt{\lambda_H}))$ where $\alpha_H^{'} \neq \alpha_H$ and we move slightly towards the direction of α_H^*. Then, the low type would strictly prefer her existing contract whereas the high type would strictly prefer the deviation. Thus, we prove that $c_H = 0$ in any separating equilibrium satisfying the intuitive criterion.

Finally, we prove that we should have $\alpha_H < \alpha_H^*$ and $\frac{T_H}{\lambda_H} = k(1+\frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}})$. Suppose not. Let $(\overline{\alpha}_H, \overline{c}_H, \overline{T}_H)$ denote the contract offer by the high type in Proposition 7. That is, $\overline{\alpha}_H$ is defined as follows:

$$\overline{\alpha}_H := \min\left\{\alpha : (1-\alpha)\lambda_L m(\Delta^{-1}(\alpha\sqrt{\lambda_L})) - k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha\sqrt{\lambda_H})) - \pi_L(\alpha_L, c_L, T_L) = 0\right\}, \tag{68}$$

where the expression

$$(1-\alpha)\lambda_L m(\Delta^{-1}(\alpha\sqrt{\lambda_L})) - k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha\sqrt{\lambda_H})) - \pi_L(\alpha_L, c_L, T_L),$$

is positive at $\alpha = \alpha_H^*$ and is strictly decreasing as α decreases. To see this, note that we can write

$$\begin{aligned} & (1-\alpha)\lambda_L m(\Delta^{-1}(\alpha\sqrt{\lambda_L})) - k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha\sqrt{\lambda_H})) \\ = & \left[(1-\alpha)\lambda_L m(\Delta^{-1}(\alpha\sqrt{\lambda_L})) - (1-\alpha)\lambda_L m(\Delta^{-1}(\alpha\sqrt{\lambda_H}))\right] \\ & + \left[(1-\alpha)\lambda_L m(\Delta^{-1}(\alpha\sqrt{\lambda_H})) - k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha\sqrt{\lambda_H}))\right], \end{aligned}$$

where $(1-\alpha)\lambda_L m(\Delta^{-1}(\alpha\sqrt{\lambda_L})) - (1-\alpha)\lambda_L m(\Delta^{-1}(\alpha\sqrt{\lambda_H})) > 0$ for all $\alpha \in (0,1)$ and is decreasing as α decreases by Assumption 2, whereas the second term

$$\left[(1-\alpha)\lambda_L m(\Delta^{-1}(\alpha\sqrt{\lambda_H})) - k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha\sqrt{\lambda_H}))\right]$$

achieves its maximum at an abandonment probability $\widetilde{\alpha} > \alpha_H^*$ and, hence, is strictly decreasing as α decreases for $\alpha \leq \alpha_H^*$. Thus, there exists a unique level of $\overline{\alpha}_H$ such that (68) holds and we have $\overline{\alpha}_H < \alpha_H^*$. This also proves that for $\alpha > \overline{\alpha}_H$, the deviation profit of the low type is strictly decreasing as α decreases, a property we will make use in the remainder of the proof.

By the second part of the proof, we need to consider only the deviations by the high type since we already know that $\alpha_L = \alpha_L^*$ and $(1-\alpha_L^*)c_L + \frac{T_L}{\lambda_L} = k(1+\frac{\Delta^{-1}(\alpha_L^*\sqrt{\lambda_L})}{\sqrt{\lambda_L}})$. Then, observe that if $\alpha_H > \overline{\alpha}_H$ and $\frac{T_H}{\lambda_H} = k(1+\frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}})$, then the low type will prefer (α_H, c_H, T_H) with c_H over (α_L, c_L, T_L). If $\alpha_H > \overline{\alpha}_H$ and $\frac{T_H}{\lambda_H} > k(1+\frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}})$, then the contract $(\alpha_H, c_H, T_H - \epsilon)$ for ϵ small enough yields strictly better profits for the high type and we would have $\pi_L(\alpha_H, c_H, T_H - \epsilon) < \pi_L(\alpha_L, c_L, T_L)$ if $\pi_L(\alpha_H, c_H, T_H) < \pi_L(\alpha_L, c_L, T_L)$ already. So suppose $\pi_L(\alpha_H, c_H, T_H) = \pi_L(\alpha_L, c_L, T_L)$. Then, by a similar argument to Lemma there exists a deviation (α, c, T) for the high type such that $\pi_H(\alpha_H, c_H, T_H) < \pi_H(\alpha, c, T)$, $\pi_L(\alpha_H, c_H, T_H) > \pi_L(\alpha, c, T)$ and $T + (1-\alpha)c\lambda_H - k(\lambda_H + \sqrt{\lambda_H}\Delta^{-1}(\alpha\sqrt{\lambda_H})) \geq 0$. Thus, we have proved that we cannot have $\alpha_H > \overline{\alpha}_H$.

Finally, we show that we cannot have $\alpha_H < \overline{\alpha}_H$. If $\alpha_H < \overline{\alpha}_H$ and $\frac{T_H}{\lambda_H} = k(1+\frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}})$, then the high type can increase α_H slightly and set $\frac{T_H}{\lambda_H} = k(1+\frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}})$, which will yield strictly better profits for her while the low type would strictly prefer $\alpha_L = \alpha_L^*$ and $\frac{T_L}{\lambda_L} + (1-\alpha_L^*)c_L = k(1+\frac{\Delta^{-1}(\alpha_L^*\sqrt{\lambda_L})}{\sqrt{\lambda_L}})$ over the deviation of the high type since $\pi_L(\alpha_H, c_H, T_H) < \pi_L(\alpha_L, c_L, T_L)$ for $\alpha_H < \overline{\alpha}_H$ and $\frac{T_H}{\lambda_H} = k(1+\frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}})$. What remains is show that $\alpha_H < \overline{\alpha}_H$ and $\frac{T_H}{\lambda_H} > k(1+\frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}})$ cannot be an equilibrium outcome in a separating equilibrium satisfying the intuitive criterion. If $\pi_L(\alpha_H, c_H, T_H) < \pi_L(\alpha_L, c_L, T_L)$, then $(\alpha_H, c_H, T_H - \epsilon)$ for ϵ small enough is a deviation for the high type that violates the intuitive criterion. Finally, observe that we cannot have $\pi_L(\alpha_H, c_H, T_H) = \pi_L(\alpha_L, c_L, T_L)$ if $\alpha_H < \overline{\alpha}_H$ and $\frac{T_H}{\lambda_H} > k(1+\frac{\Delta^{-1}(\alpha_H\sqrt{\lambda_H})}{\sqrt{\lambda_H}})$ since $\pi_L(\overline{\alpha}_H, \overline{c}_H, \overline{T}_H) = \pi_L(\alpha_L, c_L, T_L)$. ■