DEPARTMENT OF FOOD AND RESOURCE ECONOMICS
UNIVERSITY OF COPENHAGEN

IFRO Working Paper

The Hidden Cost of Specialization

Fabio Landini
Antonio Nicolò
Marco Piovesan

2013 / 9

IFRO Working Paper 2013 / 9

The Hidden Cost of Specialization

Authors: Fabio Landini, Antonio Nicolò, Marco Piovesan

Department of Food and Resource Economics (IFRO)
University of Copenhagen
Rolighedsvej 25
DK 1958 Frederiksberg DENMARK
www.ifro.ku.dk

The Hidden Cost of Specialization

Fabio Landini
MEDAlics and CRIOS, Bocconi University

Antonio Nicolò
University of Padua

Marco Piovesan
University of Copenhagen

Abstract

Given the advantages of specialization, employers encourage their employees to acquire distinct expertise to better satisfy clients' needs. However, when the client is unaware of the employees' expertise and cannot be sorted out to the most competent employee by means of a gatekeeper, a mismatch can occur. In this paper we attempt to identify the optimal condition so an employer can eliminate this mismatch and offer a team bonus that provides the first-contacted employee with an incentive to refer the client to the correct expert. We show that the profitability of this *referral contract* increases with the agents' degree of specialization and decreases with the clients' competence at identifying the correct expert. Interestingly, a referral contract may be more profitable than an individual contract -that does not pay a team bonus- even if the former provides less incentive to the agents to improve their expertise. Thus, we provide a new rationale for the use of team bonuses even when the production function depends on a single employee's effort.

Keywords: Team and Individual Contracts, Matching Client-Expert, Incentives to Refer.

JEL codes: C72; D01; D21; D86; M52.

1. Introduction

In the last decade, labor has become more complex and knowledge-intensive. As a result, employers have encouraged employees to acquire distinct and specialized expertise to be able to better satisfy specific clients' needs. For example, if we consider legal services; we all know how intricate the law can be, but only a few of us are aware of the huge number of fields and subfields in which a lawyer can specialize. The website HG.org lists 141 areas of different law practice worldwide, while at LexisNexis -a law firm locator- it is possible to search within 64 specialty and 726 sub-specialty law areas. This specialization (or hyper-specialization according to Malone et al., 2011) undoubtedly has advantages in terms of quality, speed and efficiency, since only an *expert* employee can provide an excellent service which modern clients require. However, specialization within a firm requires the employer to integrate and coordinate her highly specialized employees to make sure that *the correct expert* solves the client's needs.

Specialization within a firm in fact has a "dark side": different areas of expertise make it more difficult for the client to identify the correct expert and this increases the probability of a mismatch occurring. Clients may not have the competence to evaluate their own need or, more often, they may be unaware of the distinct expertise of each employee. Since these mismatches can be costly, not only for the client, but also for the firm – as the client only pays the firm in successful cases - employer should be prepared to deal with what we call *the hidden cost of specialization*. One way of avoiding this mismatch is to have a gatekeeper who directs the client to the correct expert. But, the gatekeeper himself may be only partially competent and may not eliminate the possibility of a mismatch. Moreover, in many organizations each employee is directly linked with a set of clients and there is not the possibility to centralize the matching between clients and employees in order to sort out each client to the expert who better answers to her needs. Financial experts, lawyers, advisors working in law and consulting firms are often directly contacted by clients through their social networks. In this paper, thus, we focus on those cases where the organizational solution of centralizing the matching between clients and expert using gatekeepers is unfeasible, and therefore the employer can only design a contract that induces agents to cooperate and direct the client to the expert who has the right competence to satisfy her needs. The employer can avoid a mismatch by using a (costly) team bonus that in case of a mismatch provides the first-contacted agent with an incentive to refer the client to the correct expert, i.e. the one who is more likely to successfully serve the client. This feature resembles many real life situations. For instance, if we consider specialized professionals such as lawyers or consultants, but also shop assistants, travel agents; in all these examples the client may contact the wrong employee which will reduce the probability of being served successfully with a resulting loss in profit for the employer.

We propose a novel explanation on why in some organizations we observe compensation schemes with a team bonus, even if the production function has not any team component. Cooperation among agents does not directly affect workers' productivity, but indirectly by favoring an efficient matching among employees' expertise and clients' needs.

In our model, we assume there exists asymmetric information between the client and the agents (employees) working for a principal (employer). In other words, we are open to the possibility that the client is not fully able to identify the expert for her need. The principal has therefore to evaluate the expected cost of this mismatch and compare it with the cost of a contract that induces the non-expert to refer the client to the expert. In fact, a specific feature of our model is that the client's incompetence decreases the probability of the agent successfully serving the client and therefore the principal's expected profits.

In our analysis, we consider different types of contracts and analyze how these contracts change agents' behavior. If the principal offers an Individual Contract (IC) that pays a bonus (only) to the agent who successfully serves the client, the first-contacted agent has no incentive to refer the client to another agent even if this agent has more expertise. The first-contacted agent will always try to serve the client, no matter what. Instead, if the principal offers a Referral Contract (RC) that rewards all the agents if one of them successfully serves the client, then the first-contacted agent has an incentive to refer the client to the agent who has the greatest expertise. However, RC is more expensive as the principal has to pay a team bonus to *all* the agents even if they do not serve the client. Our setting allows us to analyze the trade-off between an incentive to exert effort (or to invest in expertise) and an incentive to refer (or to avoid the mismatch). We analyze this trade-off between IC and RC and we identify the conditions under which one contract is more profitable than the other.

By offering an RC, the principal rewards each agent for his work as a *gatekeeper*. However, *ex ante* our agents are identical and therefore each agent can be the expert or not depending on the client's specific need. This peculiarity distinguishes our work from Shumsky and Pinker (2003) who examine services in which customers encounter a gatekeeper (such as a physician) who makes an initial costly diagnosis of the customer's problem and then may refer the customer to a specialist. The gatekeeper may also attempt to solve the problem, but the probability of treatment success decreases as the problem's complexity increases. In our model, there are not gatekeepers and each agent may be the physician or the specialist with the same probability: Nature, clients' needs and agents' expertise determine their role. Therefore, instead of studying the optimal referral rate in the manner of Shumsky and Pinker (2003), we study the optimal contract that the principal has to offer.

To explain in more detail; in this paper we present a model with a principal, *n* agents and a client who is randomly matched with one of the agents. Our model includes sequential moves in three stages. In the first stage, the principal chooses a contract: IC or RC. In the second stage, having observed which contract the principal choses, agents select the level of effort they want to make to increase their expertise. In the third stage, Nature determines the client's need and consequently the fitness of the agents' expertise given this need. The first contacted agent can either serve the client, or if he is not the expert, refer the client to the expert. If he serves the game ends; if he refers, then the expert serves the client and the game ends.

A distinctive feature of our model is that the probability that the agent will successfully serve the client does not only depend on their investment in expertise. This probability has a second component, i.e., the goodness of fit which is determined by Nature. This stochastic component is not under the control of the agents or the principal. It depends on the ability of the client to address the most competent employee and it is represented in our model by the goodness of the fit for the first mover. A possible mistake by the client can therefore create a mismatch between her needs and the agents' expertise and therefore a loss of profit (in case of an IC) or additional costs (in case of a RC) for the principal. In addition, our agents are ex-ante identical and they therefore all have the same probability of being the expert.

We solve our model by backward induction and in separate cases. First we analyze the case with completely informed clients (first best solution). In this case the client contacts the correct expert and therefore there is no need for the principal to use a referral contract. Then we extend this trivial case to allow for some asymmetric information between the client and the agents. We compute the optimal action and effort if the principal offers his agents an IC and a RC. Then we determine the condition under which a referral contract is more profitable than an individual contract. Counter intuitively, our model shows that an RC can be more profitable than an IC even if the former provides less incentive for the agents to increase their expertise. Thus, we provide a new rational for the use of team bonuses even when the production function depends on a single employee's effort.

Finally, we extend our model to the case of multiple levels of expertise and we introduce an Imperfect Referral Contract (IRC) in which the principal does not exclude the possibility of a mismatch, but he only eliminates the most costly mismatches. We show under which condition this contract is better than the other two and *vice versa*. Our numerical example shows which parameters determine the relative profitability of these contracts and help the reader understand the mechanisms that determine the hidden cost of specialization. We conclude our paper by stressing the importance for the employer to consider the dark side of specialization and we suggest possible alternative interventions to make the mismatch less likely.

2. Literature Review

Our paper studies the role of team incentives in firms and provides a new rational for the use of team bonuses even when the production function depends on a single employee's effort. The previous literature (Hamilton et al. 2003; Boning et al. 2001) provides empirical evidence suggesting that, in spite of the free-riding problems suggested by the seminal paper by Alchian and Demsetz (1972), team incentives increase productivity. Such an effect tends to be greater, the greater the heterogeneity of the agents and the complexity of the production lines. One explanation for this result stems from the literature on side contracting in principal-multi-agents relationships, where it is shown that, under certain conditions, the free-riding problem under work teams can be resolved via "mutual help", cooperation and peer pressure (Itoh, 1990; Itoh, 1991; Itoh, 1992; Itoh, 1993; Macho-Stadler and Pérez-Castrillo, 1993). In this respect, our paper provides an alternative explanation: despite the free-riding problem, team incentives may in fact increase firm productivity by eliminating the mismatch between clients' needs and employees' expertise. Levin and Tadelis (2005) show that profit-sharing partnerships alleviate the problem of asymmetric information regarding a client's knowledge concerning the ability of the employees, because profit-sharing partnerships tend to be more selective when hiring new employees than profit-maximizing firms. They predict that partnerships will emerge when human capital has a central role in determining product quality and when clients are in a disadvantageous position with respect to firms when assessing the ability of employees. We predict that compensation schemes where the gains of a successful service are shared equally among the employees emerge when human capital specialization is important and when clients find it hard to recognize employees' specific expertise.

Our paper contributes to the literature on the matching (and coordination) of tasks (or clients) and agents within different institutional and firm settings (see Garicano 2000; Garicano and Santos 2004; Garicano and Hubbard, 2009; Inderst and Ottaviani 2009; Park 2005; Xu 2011; Blouin and Bourgeon 2008; Fuchs and Garicano 2010; Kinshuk Jerath and Zhang 2010; Corts 2007; Faraj and Sproull, 2000; Harris and Raviv, 2002; Epstein et al., 2010). As in the paper by Inderst and Ottaviani (2009), in our model the conflict of interest between the client and the agent arises endogenously from the agent's compensation scheme, but while Inderst and Ottaviani consider the mismatching between the product characteristics and the customer's preferences, we analyze the effects of a mismatch between the agents' expertise and the client's need. Moreover, we make a restriction in that the principal only observes which agent serves the client without receiving any ex-post signal regarding the agent's misbehavior.

The closest paper to ours is the seminal paper by Garicano and Santos (2004) in which the authors analyze the problem of matching opportunities with talent. However, our model differs from theirs in many respects. Firstly, Garicano and Santos analyze which type of contract among experts generates an optimal allocation of clients according to the experts' talents, while we look at the mismatching problem within an organization and we study the optimal contract that a employer should propose to employees with different expertise to induce cooperation among them. Secondly, Garicano and Santos study the matching of opportunities with talent when costly diagnosis confers an informational advantage to the agent who undertakes it. Therefore, the problem of moral hazard results from the lack of information regarding the value of the opportunity for the agent who (eventually) is referred. In our setting, agents select the level of effort that determines their degree of expertise prior to the matching and the value of the opportunity is common knowledge among them. The moral hazard problem may eventually arise because of the incentive to free-ride on the expertise of others. Finally Garicano and Santos look at the case when the clients are completely uninformed and are randomly matched with the expert, while we are interested in studying how the optimal contract varies depending on the degree of clients' knowledge.

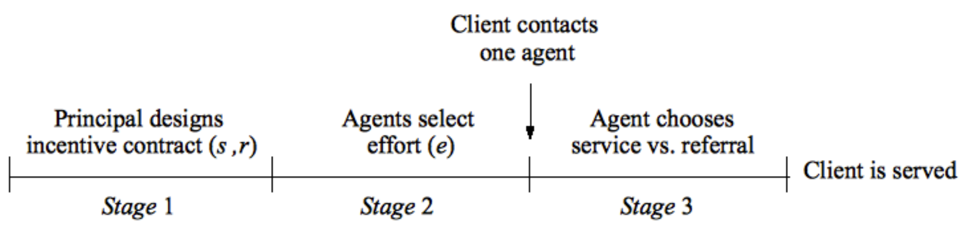
Shumsky and Pinker (2003) suggest solving the mismatching problem within a firm by introducing a gatekeeper with the task of evaluating the client's need and directing the client to the correct expert. We depart from this literature for two reasons. Firstly, our analysis considers organizations in which it is not possible to have gatekeepers because each agent is directly linked with a set of clients. Secondly, we analyze the cost of dealing with different expertise and giving the correct incentive to refer; this cost can be lower than the cost of hiring a competent gatekeeper and therefore its presence is inefficient.

Finally, it is worthy mentioning that in our model the asymmetric information between agents and client produces a cost for the client and the firm. This cost of asymmetric information, as in Inderst and Ottaviani (2009), represents the flip-side side of the coin in terms of the literature on credence goods (Darby and Karni, 1973; Debo and Toktay, 2008; Dulleck et al., 2011). In fact, credence goods produce extra costs for clients, but extra profits for the seller (for instance unnecessary treatments, over-prescribing and needless repairs).

3. Model

We consider a stylized model in which one client has a specific need to satisfy. This need can be a service that has to be provided or a good that has to be sold. The client contacts one of the n (> 1) agents (*employees*) working for a principal (*employer*). The agent can successfully serve the client or not. In the former case, the client pays x_S to the principal; in the latter, the client pays x_F. For the sake of simplicity we parameterize x_F =0 and we denote $x_S = x$. Figure 1 depicts the timing of our model. In Stage 1, the principal specifies an incentive contract for the agents. In Stage 2, each agent i (for $i = 1,\ldots, n$) decides simultaneously how much effort to exert. This effort $e_i \in(0,1)$ has a cost for the agent, $c(e)=\theta e^2/2$, where $\theta > 0$. In Stage 3, the client contacts one agent who has the opportunity to either serve the client, or refer the client to another agent within the firm.

Figure 1 – Timing



Note. The letters in parentheses are the variables selected at each stage.

Each of the n agents in the firm has a different specialization and it is the specific need of the client that determines who is the correct expert. Before the matching with the client, all agents share the same expectations regarding the probability of being an expert (they are *ex-ante* identical). Nature randomly selects one of n client's needs from a uniform probability distribution. For any agent i we define $\lambda_i \in[0,1]$ as the fitness of i's specialization with respect to the client's need: if i is *the expert*, then $\lambda_i = 1$; if i is a *non-expert*, then $0 \leq \lambda_i \leq 1$. We assume here that all non-experts have the same λ; this assumption will be removed in Section 6. The difference in specialization between the expert and the non-expert is therefore $(1 - \lambda)$.[1]

[1] Notice that in this setting we model specialization by taking as given the heterogeneity of customers' needs, as captured by parameter n (i.e. the number of different types of agent's specialization available in the firm). An alternative could be to model a reduction in specialization by assuming that there exist $k < n$ agents, each with a

In our model the probability of successfully serving the client has two components: one is the fitness between the client's need and the agent's specialization and the other is the effort exerted in Stage 2. This probability is equal to $\lambda_i e_i^{\beta}$, where $0 \leq \beta < 2$ is a parameter that captures the effect of effort on the probability of success. We assume this parameter is constant among the agents within the firm and depends on the specific sector in which the firm operates.

In Stage 3, the expert is matched with the client with probability $p \in (0,1)$. This probability can be seen as the client's competence in identifying the correct expert or similarly the competence of the gatekeeper of the firm in identifying the agent who better fits the client's need. In all the cases in which the non-expert agent is matched with the client, we say that a *mismatch* occurs.

The expected cost of mismatching can be thus defined as:

$$\mu(\lambda, p) = \frac{(n-1)(1-p)}{n}(1-\lambda), \tag{1}$$

where the first term is the probability of a mismatch occurring, while the second term is the size of this mismatch; $(1 - \lambda)$ is a measure of the degree of the agents' specialization (or overlap of specialization) that exists within the firm. When $(1 - \lambda)$ is small, experts and non-experts have similar specializations (high overlap) and the size of the mismatch is low. In contrast, when $(1 - \lambda)$ is large, the agents are hyper-specialized (low overlap) and the size of the mismatch is large. The previous literature assumes $p = 1$, i.e. that the client is always able to identify the expert with a probability of one, and neglects the case with $p < 1$. In this paper we consider the cases in which p (and λ) are lower than 1, and therefore the expected cost of mismatching for the principal is greater than zero.

In our model, the principal observes whether the client has been successfully served or not but he cannot observe the client's need or identify the expert who best fits the client's needs, or the identity of the agent who was first matched with the client. In other words, the principal cannot observe whether referral has occurred. This implies that the principal cannot offer an incentive to the agent who makes a referral. The act of referral is difficult to rule in a contract and/or too costly to verify. Each agent knows the other agents' specialization but only the agent who matches the client

different type of specialization. This way of doing, however, would force us to introduce additional assumptions on the distribution of the agents' expertise and would not affect the main insights of the model.

discovers the client's specific need. Finally, the effort chosen in Stage 2 by agent i is neither verifiable nor observable by the principal or any other agent $j \neq i$.

As previously mentioned, in Stage 1 the principal specifies an incentive contract. We define an incentive contract as pair (s,t), where s (≥ 0) is an individual bonus paid to the agent who successfully serves the client and t (≥ 0) is a team bonus paid to other $n - 1$ members of the firm. In our analysis we compare two types of contract: an *individual contract* where $s > 0$ and $t = 0$ and a *referral contract*, where both $s > 0$ and $t > 0$. A trade-off for the principal exists between these two contracts. An individual contract may produce a mismatch since the first agent with whom the client is matched always serves the client, even if he is not the expert. A referral contract may alleviate the mismatching problem but it is more costly. In the rest of the paper we compare the profitability of these two contracts within the (λ, p) space.

Before proceeding with the analysis, we want to underline some of the assumptions we made to keep our model simple and tractable. Firstly, all agents have been hired and contracts cannot be rejected. Secondly, the principal and the agents are risk neutral. Thirdly, we assume that there is no effort exerted by the agent who serves the client. This simplifying assumption does not affect our main results. Since the principal knows which agent serves the client, any moral hazard problem that may arise when ex-post effort is not verifiable can be easily solved by the usual incentive scheme paid to the agent who serves the client.[2]

4. The Principal-Agent Problem with two Levels of Specialization

In this section, we present our results. In Section 4.1, we determine the first best contract with a perfectly informed client, while in Section 4.2, we solve the model for the individual and referral contract when some information asymmetry between the agent and client exists. We then compare the two solutions and find the optimal contract. We solve our model by backward induction. First, we study the agent's decision between serving and referring in Stage 3 and we determine under which conditions referral occurs. Then, we study the effort choice problem in Stage 2 and compute the agents' best response function. Finally, we solve the optimal bonus problem faced by the principal in Stage 1. In solving the model, we focus on symmetric equilibria.

[2] e.g. lawyers' billable hours.

4.1 Solution for the first-best contract with a perfectly informed client (p = 1)

Firstly we solve the model under the assumption that the client is perfectly informed. In this case the analysis is trivial: when $p = 1$ a mismatch between the client's need and the agent's specialization never occurs since the client always contacts the correct expert first. Thus, the optimal contract is an individual contract where the bonus is only paid to the expert. Since this case represents our benchmark, we determine the agent's optimal level of effort and the profit for the principal.

LEMMA 1: *When the client is perfectly informed about the agents' specialization, the optimal contract pays a bonus equal to* $s^* = \frac{x\beta}{2}$ *to the agent who successfully serves the client and induces an effort level equal to* $e^* = \left(\frac{\beta^2 x}{2\theta n}\right)^{\frac{1}{2-\beta}}$. *The principal obtains a level of profit equal to*

$$\pi^* = \left(\frac{\beta^2 x}{2\theta n}\right)^{\frac{\beta}{2-\beta}} \frac{x(2-\beta)}{2}.$$

Proof: see the Appendix

Clearly when the client has perfect knowledge of the agents' specialization, the agent who serves the client will always be an expert and the fitness of the non-expert λ becomes irrelevant.

4.2 Solution for the individual and referral contracts with an imperfectly informed client (p < 1)

It can be difficult for a client to identify the correct expert as she may be unaware of the employees' areas of specialization or she may simply contact an agent she knows without considering other agents who may have specialized knowledge that is more relevant to her need. In these cases a mismatch can occur. What are the consequences for the agents and the principal?

The (first-contacted) agent's decision in Stage 3 depends on two components: the fitness (λ) between the agent's specialization and the client's need and the size of the bonuses (s and t) specified by the principal. If the first-contacted agent is the expert, it is optimal for him to serve as shown above. In contrast, if he is not the expert, he will prefer to refer if and only if:

$$\lambda e^\beta s \leq \underline{e}^\beta t \tag{2}$$

where the left-hand side represents the expected return when the agent serves the client, and the right-hand side represents the expected return when the agent refers the client to the expert (in this case $\underline{e}$ is the effort exerted by the expert). In a symmetric equilibrium Eq. (2) reduces to:

$$t \geq \lambda s,$$

that is the agent refers whenever the team bonus is higher than the expected value of serving the client and receiving s. Depending on the contract offered by the principal (individual vs. referral), the decision of the first-contacted agent will differ at this stage.

Individual contract

We first consider the individual contracts according to which the principal pays a bonus s to the agent who successfully serves the client, and does not pay any team bonus, *i.e.*, $t = 0$. In this case, the first agent to be matched with the client always serves. The following lemma describes the main features of the optimal individual contract.

LEMMA 2: *If the principal offers an individual contract* $(s,0)$ *and* $p < 1$*, the optimal individual contract pays a bonus* $s^I = \frac{x\beta}{2}$ *to the agent who successfully serves the client and induces an effort equal to* $e^I = \left\{\frac{\beta^2 x}{2\theta n}\left[p + (1-p)\frac{1+(n-1)\lambda}{n}\right]\right\}^{\frac{1}{2-\beta}}$. *The principal obtains a level of profit equal to*

$$\pi^I = \left(\frac{\beta^2 x}{2\theta n}\right)^{\frac{\beta}{2-\beta}}\left[p + (1-p)\frac{1+(n-1)\lambda}{n}\right]^{\frac{2}{2-\beta}}\frac{x(2-\beta)}{2}.$$

Proof: see the Appendix

The level of effort exerted in the individual contract when $p < 1$ is lower than the level of effort e^* exerted in the first-best and coincides with e^* only when $p = 1$. The marginal benefit of effort is lower when mismatching occurs. At Stage 2, when the agent decides how much effort to exert, each agent serves the client with probability $1/n$, regardless of the client's competence p. However, conditional on being the expert, in the first-best case, the agent is matched with the client with probability one, while in case $p < 1$, the agent is matched with the client only with probability p + (1

$-\ p)\ 1/n$. Conditional on not being the expert, in the first-best case, the agent is matched with the client with probability zero, while in case $p < 1$ with probability $(1-p)1/n$. Since the marginal productivity of the effort is lower when the agent who serves the client is not the expert, the possibility of mismatching reduces the agents' incentive to invest in their own specialization. It follows that the principal's profit is reduced for two reasons: the direct negative effect of mismatching on the probability of success which is lowered by $(1-\lambda)$ when the agent is not the expert, and the indirect effect due to the reduction in the effort exerted by each agent.

Referral contract

We now consider a contract that can solve the mismatching problem, *i.e.* a contract that induces the first-contacted agent to refer the client every time he is not the expert. We assume here that the expert always serves the client. In Section 6 we discuss the possibility that an agent may decide to free-ride within the organization and to shirk at Stage 2 without investing any effort. An agent who shirks always refers the client to another agent regardless of his specialization.

Lemma 3 summarizes the main feature of the optimal referral contract with no shirking:

LEMMA 3: *If the principal offers a referral contract* (s,t) *and* $p < 1$*, the optimal referral contract pays a bonus* $s^R = \dfrac{\beta x}{2[1+(n-1)\lambda]}$ *to the agent who successfully serves the client, a team bonus equal to* $t^R = \dfrac{\beta x\lambda}{2[1+(n-1)\lambda]}$ *to all the other agents, and induces an effort equal to* $e^R = \left\{\dfrac{\beta^2 x}{2n\theta}\left[\dfrac{1}{[1+(n-1)\lambda]}\right]\right\}^{\frac{1}{2-\beta}}$. *The principal obtains a level of profit equal to* $\pi^R = \left(\dfrac{\beta^2 x}{2n\theta[1+(n-1)\lambda]}\right)^{\frac{\beta}{2-\beta}} \dfrac{x(2-\beta)}{2}$.

Proof: see the Appendix.

Two remarks. First, it is easy to check that the value of the team bonus t does not affect the incentive to exert effort and that it is therefore optimally fixed at the minimum level to induce referral, $t = \lambda s$. Second, the incentive to exert effort is provided by the individual bonus s^R, which is equal to $s^* / [1+(n-1)\lambda]$. If $\lambda = 1$ there is no specialization because each agent is equally good at solving the client's need: paying a team bonus would only reduce the incentive to exert effort because the monetary incentive would be spread among all the agents instead of only being paid to the agent who successfully served the client. If $\lambda = 0$ there is an extreme degree of specialization and a non-expert has no chance of successfully serving the client. In this case, the referral contract mimics the first-best contract because a non-expert has no incentive to serve the client: the team bonus goes to zero and all the monetary incentives are concentrated in the individual bonus s^R. It follows that the level of effort induced by the referral contract is decreasing in λ.

Comparison between individual and referral contract

Comparing the optimal bonuses, effort and profit under the two contracts, as well as in the first-best, we derive the following results:

PROPOSITION 1: *For any λ, p, and β, $s^* = s^I = s^R + (n-1)\, t^R$ i.e. the total amount of monetary incentives is the same under the three contracts.*

Proof: It follows directly by Lemmata 1, 2 and 3.

The intuition behind Proposition 1 is as follows. In the first-best, as well as in the individual contract, effort is the only component of the probability of success that the principal has to incentivize. In the former there is no mismatch, in the latter a mismatch can occur but the principal cannot avoid it. Therefore, in both contracts, the principal allocates all his resources to incentivize effort. This is not the case under a referral contract as the principal can solve the mismatching problem by specifying t. However, the resources allocated to avoid the mismatching have to be subtracted from the resources used to induce effort. The expected cost of mismatching determines whether it is profitable for the principal to solve the mismatch and, consequently, which contract maximizes his profits.

PROPOSITION 2:

a) If $\lambda < \lambda^e$ *and* $p < p^e(\lambda)$, *where*

$$\lambda^e = \frac{n^{\frac{1}{2}} - 1}{n-1} \qquad p^e(\lambda) = \frac{n - [1+(n-1)\lambda]^2}{n[1+(n-1)\lambda] - [1+(n-1)\lambda]^2},$$

then the optimal referral contract induces more effort than the optimal individual contract.

b) If $\lambda < \lambda^\pi$ *and* $p < p^\pi(\lambda)$, *where*

$$\lambda^\pi = \frac{n^{\frac{2}{2+\beta}} - 1}{n-1} \qquad p^\pi(\lambda) = \frac{n - [1+(n-1)\lambda]^{\frac{2+\beta}{2}}}{n[1+(n-1)\lambda]^{\frac{\beta}{2}} - [1+(n-1)\lambda]^{\frac{2+\beta}{2}}},$$

then the optimal referral contract is more profitable than the optimal individual contract.

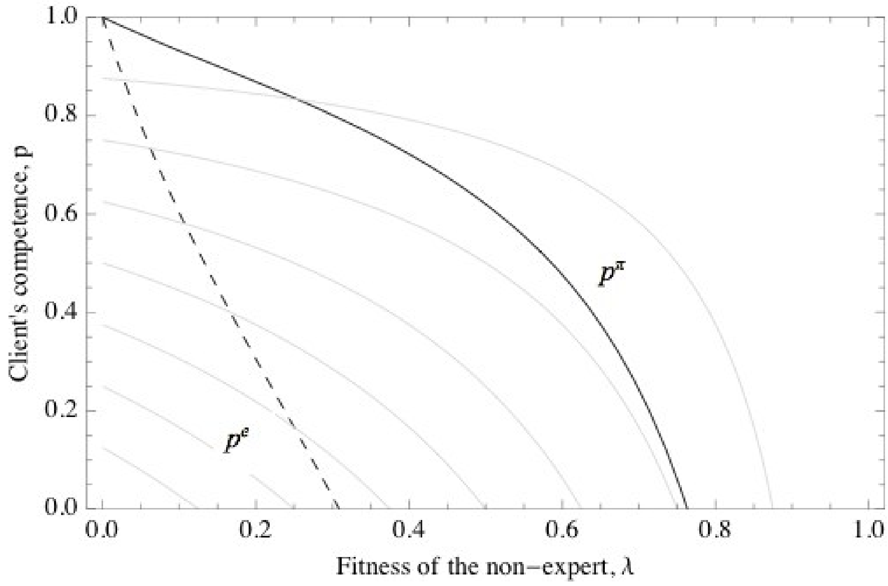
Proof: see the Appendix.

COROLLARY: *Pairs* (λ, p) *exist such that a referral contract is more profitable than an individual contract even if it induces less effort. The reverse is not true.*

Proof: It follows directly from the fact that $\lambda^e < \lambda^\pi$, and $p^e(\lambda) \leq p^\pi(\lambda)$ for any $0 < \lambda < 1$ and $0 \leq \beta < 2$.

Figure 2 reports the content of Proposition 2 in the (λ, p) space discussed above. The dotted line corresponds to p^e, whereas the continuous line corresponds to p^π. The grey lines represent the level curves associated with the expected cost of mismatching $\mu(\lambda, p)$ defined in Eq. (1), where such cost rises moving towards the south-west corner of the figure. Point a) in the proposition states that for all points below the line p^e a referral contract induces more effort than an individual contract. At first glance, this result seems counter-intuitive. However, when both λ and p are low, agents are highly specialized and the client is unable to identify the agent who best fits her needs. In this case the mismatching is severe and therefore the agents have little incentive to exert effort under an individual contract. In fact, their effort would pay out only in the unlikely event that an agent is the expert and is contacted by the client, whereas under a referral contract, effort has a higher value since it always pays out when the agent is the expert – both when the client contact directly him or the client is referred to him. Therefore, for sufficiently low values of λ and *p,* a referral contract can induce more effort than an individual contract.

Figure 2 – Contract thresholds



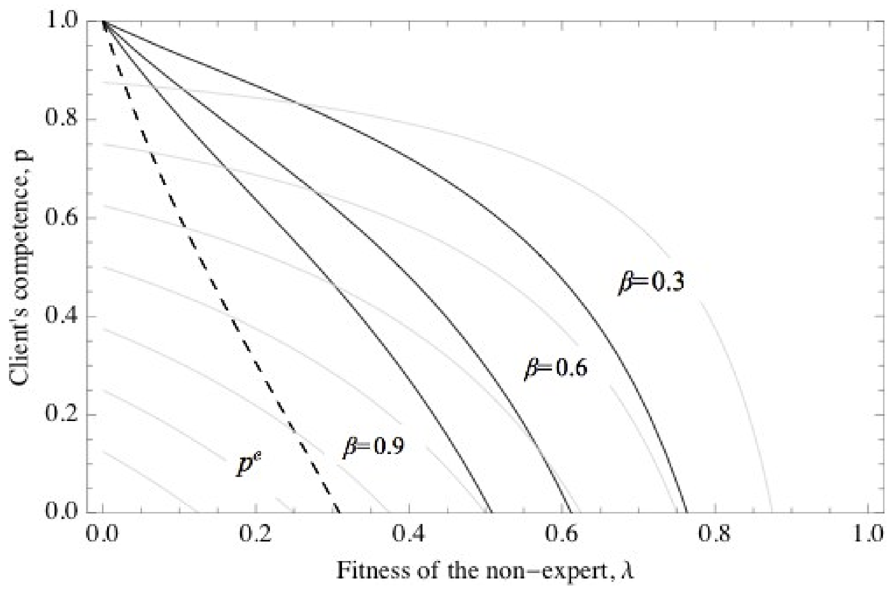
Note. p^e and p^π are computed assuming that $n = 5$, $\beta = 0.3$. For all points below p^e (p^π) an optimal referral contract induces more effort (is more profitable) than an optimal individual contract. For all points between p^e and p^π the optimal referral contract is more profitable despite inducing lower effort.

In addition to the agents' effort, λ and p affect the profitability of these two contracts. The second part of Proposition 2 claims that for all points below the curve p^π a referral contract is more profitable than an individual contract. In fact, for low values of λ and *p,* the expected cost of mismatching is high and this negatively affects the principal's profit. As we saw in Proposition 1, the principal faces a trade-off: avoid the mismatch or induce effort? Thus, for sufficiently low values of λ and *p,* the elimination of mismatching has benefits that more than compensate for a lower level of effort and a referral contract is more profitable than an individual contract.

Interestingly, we find that by considering the first and second part of Proposition 2 together, it is possible to identify a whole set of points in the (λ, p) space for which there is no monotonic relationship between effort and profit. At these points, a referral contract is more profitable than an individual contract, despite inducing lower effort. The reverse, however, does not hold. This is due to the fact that, in our setting, effort is only one of the factors which affect the principal's expected profit. Since only a referral contract can solve the mismatch, this contract may be more profitable even if it presents higher costs to induce effort. When information asymmetries between clients and agents exist, contracts should be designed not only to increase the level of effort, but also to reduce the possibility of a mismatch.

The extent to which a referral contract performs better than an individual contract (and *vice versa*) also depends on the (relative) importance of effort in the probability of success (β). A large β reflects situations in which the effort exerted before the match is highly relevant in determining the probability of success, such as when a lawyer is required to study new laws to keep his specialization up to date. In these cases, we would expect an individual contract to perform better than a referral contact in a larger set of the (λ, p) space. On the contrary, when β is small, the relevance of effort is limited. Figure 3 confirms this intuition: as β increases, p^{π} tends to get closer to threshold p^e and the region in which an individual contract does better than a referral contract increases.

Figure 3 – Contract thresholds when β changes



Note. p^e *and* p^π *are computed assuming* $n = 5$. *Each continuous line represents* p^π *for different levels of* β. *As* β *increases, the set of points where the optimal referral contract is more profitable than the optimal individual contract reduces.*

Propositions 1 and 2 have important consequences for the design of contracts in firms. In all cases in which agents' hyper-specialization reduces the capability of the client to identify the correct expert, the principal may find it profitable to use the bonus t to avoid this mismatch. The principal has to move the resources from bonus s to t and therefore to reduce the incentive to exert effort. This trade-off has been neglected in the previous literature and is what we call the *hidden costs of specialization*.

The hidden costs of specialization of a contract can be computed by taking the difference between the profit that firms can earn in the first-best case and the profit that can be earned with this contract when $p < 1$. Such a difference reflects the profit loss that the firm incurs due to the client's limited ability to identify the correct expert. Formally, the hidden costs of specialization in using an individual and referral contract, denoted respectively by χ^I and χ^R are:

$$\chi^I(\beta,\theta,n,x,p,\lambda) = \left(\frac{\beta^2 x}{2\theta n}\right)^{\frac{\beta}{2-\beta}} \frac{x(2-\beta)}{2}\left\{1-\left[p+(1-p)\frac{1+(n-1)\lambda}{n}\right]^{\frac{2}{2-\beta}}\right\} \tag{3}$$

$$\chi^R((\beta,\theta,n,x,p,\lambda)) = \left(\frac{\beta^2 x}{2\theta n}\right)^{\frac{\beta}{2-\beta}} \frac{x(2-\beta)}{2}\left\{1-\frac{1}{[1+(n-1)\lambda]^{\frac{\beta}{2-\beta}}}\right\} \tag{4}$$

Eq. (3) shows that under an individual contract, the hidden cost of specialization is associated with the risk that a mismatch occurs. In fact χ^I decreases both in λ and in p, the expected cost of mismatch. In contrast, Eq. (4) shows that under a referral contract, the hidden cost of specialization is associated with the resources necessary to avoid mismatching, i.e. the larger λ (i.e. the lower the degree of specialization), the more resources are needed since the principal needs to design a larger team bonus to induce a non-expert to refer. The extent to which one contract is more profitable than the other depends on the relative size of these two costs. As long as the cost of avoiding mismatching is lower (larger) than the expected profit loss due to mismatching, a referral contract (individual contract) will be preferred over an individual contract (referral contract).

5. The Principal-Agent Problem with Multiple Levels of Specialization

In the previous section we assumed that all agents who are not experts share the same level of specialization. Under this assumption, the principal either chooses an individual contract that never induces referral, i.e. the first agent who is matched with the client always serves; or a referral contract that always induces referral, i.e. any non-expert agent who is first matched with the client always refers to the expert. In many situations the degree of expertise is not a dichotomous variable and it can take different levels: agents' specialization may differ depending on the needs of the client, and a principal can choose under which circumstances the referral should induce an agent to refer the client to a more expert colleague. An *imperfect referral contract induces referral when the fitness with the client is low, but not when it is at an intermediate level.*

In this section we present a simple extension of the model presented in Section 4 where we consider the possibility of the principal designing an imperfect referral contract. The latter is defined as a pair (s,t) such that $s > 0$, $t > 0$ and the client is served by an agent who has a sufficiently high λ (not necessarily the expert). Intuitively, this situation does not differ significantly from the case discussed in Section 4 and the extent to which such a contract outperforms an individual contract and/or a referral contract depends on the expected cost of mismatching. If such a cost is high (low) for all levels of λ available in the firm, then a referral contract (individual contract) will probably be preferable. On the contrary, if such a cost is limited for at least some levels of λ, then an imperfect referral contract will probably be the most convenient solution.

To model this intuition, we consider the following setting. Similar to the case discussed in Section 4, any agent i can either be an expert or a non-expert. If i is an expert, then $\lambda_i = 1$. If i is not an expert, then there are two possibilities: with probability σ i can be a "mid" expert with $\lambda_i = \lambda_M$; with probability $1 - \sigma$ i can be a non-expert with $\lambda_i = \lambda_L$ (where the suffix L stands for low). This implies that we now assume $\lambda \in \{1, \lambda_M, \lambda_L\}$ with $1 \geq \lambda_M \geq \lambda_L$. All the rest of the model, timing and information structure is maintained as in the previous section. Once again we solve the model by backward induction. The complete solution of the model is presented in the Appendix. In what follows, we report the main intuition behind the results.

In this setting, the principal's optimal choice has to be based on a more detailed evaluation of expected mismatching. In particular, for any p, there are two types of mismatch that the principal should take into account. One is the mismatch due to the difference in specialization between the expert and the non-expert, *i.e.*, $(1 - \lambda_L)$. The other is the mismatch due to the difference in specialization between the expert and the mid-expert, *i.e.*, $(1 - \lambda_M)$. Depending on the size of such mismatches, as well as on the likelihood that they arise (*i.e.*, on p), the trade-off between inducing effort and avoiding mismatching will be resolved in different ways.

The principal can design three distinct contracts. In each contract a different portion of the expected mismatch is eliminated. The first contract is an individual contract with no team bonus, *i.e.*, $s > 0$ and $t = 0$. This is the same contract as the one considered in Section 4. Any agent who is matched first with the client has an incentive to serve, and the expected mismatching is maximal. The second contract is an imperfect referral contract. In this case the principal fixes a level of t (> 0) such that the mid-expert serves the client, while the non-expert has an incentive to refer. This implies that some resources have to be taken out from the individual bonus to ensure that the non-expert refers the client to the expert colleague. Using this contract, the expected mismatching associated with the difference $(1 - \lambda_L)$ is thus eliminated. Finally, the principal can design a referral contract. In this case t has to be increased even further to ensure that the mid-expert also has an incentive to refer. By using such a contract, the difference $(1 - \lambda_M)$ is also eliminated and no mismatch occurs. Obviously, we already know from Section 4 that this comes at a cost, which is related to the resources that have to be subtracted from s.

Given these different contracts and the role that they play in insuring the principal against a different proportion of mismatch, it is clear that the optimal choice depends on the expected cost of such a mismatch. In other words, similar to the case with two levels of specialization, the optimal contract depends on the value λ and p.

PROPOSITION 3:

a) For any λ_M, λ_L, p, and β, $s^* = s^I = s^R + (n-1)\,t^R = s^{IR} + (n-1)\,t^{IR}$ i.e. the total amount of monetary incentives is the same under the three contracts and is equal to the incentives paid under the first-best.

b) The profit for the principal in each contract is equal to:

$$\pi^I = \left(\frac{\beta^2 x}{2\theta n}\right)^{\frac{\beta}{2-\beta}} \left[p + (1-p)\frac{1+(n-1)\bar{\lambda}}{n}\right]^{\frac{2}{2-\beta}} \frac{x(2-\beta)}{2}$$

$$\pi^R = \left(\frac{\beta^2 x}{2n\theta\left[1+(n-1)\lambda_M\right]} \right)^{\frac{\beta}{2-\beta}} \frac{x(2-\beta)}{2}$$

$$\pi^{IR} = \left(\frac{\beta^2 x}{2n\theta\left[1+(n-1)\lambda_L\right]} \right)^{\frac{\beta}{2-\beta}} \left\{ p + (1-p)\frac{1+(n-1)\left[1-\sigma(1-\lambda_M)\right]}{n} \right\}^{\frac{2}{2-\beta}} \frac{x(2-\beta)}{2}$$

where $\overline{\lambda} = \sigma\lambda^M + (1-\sigma)\lambda^L$. *Depending on the value of* λ_M, λ_L *and p the optimal contract is* max $\in$ $\{\pi^I, \pi^R, \pi^{IR}\}$.

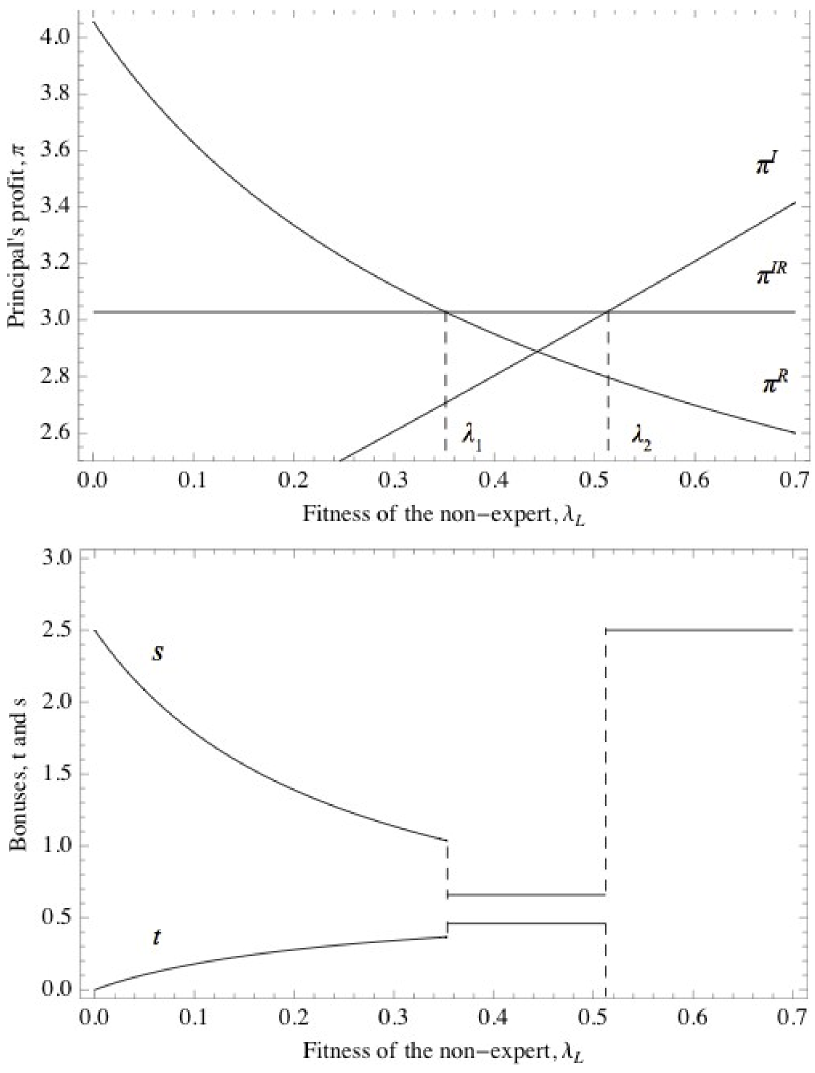
Proof: See Appendix.

To provide a full characterization of the optimal contract is not a trivial task due to the number of relevant parameters. For this reason, we present here some numerical simulations that demonstrate in an intuitive and simple way how the optimal contract varies when some parameters vary.

Figure 4 shows how the total profit of the three contracts changes with variations in the value of λ_L, taking λ_M and *p* as given. The portion of lines associated with the contract that generates the highest profit represents the frontier of the optimal contract. The panel below reports the value of the optimal bonuses *s* and *t* as the frontier of the optimal contract changes. As can be seen, a non-empty set of parameters in which each of the three contracts is optimal exists. For $\lambda_L < \lambda_1$ an imperfect referral contract is optimal. This part of the graph reflects the situation in which the fitness of the non-expert is very small compared to both the expert and the mid-expert. Hence it is profitable for the principal to spend part of his resources to insure against the risk that the non-expert serves the client. The cost of doing this is very small in the imperfect referral contract, because for λ_L close to zero it does not take a large *t* to only induce the non-expert to refer. The same is not true in the referral contract where the principal sets *t* to induce both the non-expert and the mid-expert to refer. This makes the imperfect referral contract optimal. As λ_L increases, however, the size of *t* under imperfect referral increases too and consequently (as suggested by point a) in Proposition 3) the size of *s* reduces (notice that in the figure we assume that *n* = 5 so the total size of incentives is given by multiplying *t* by four and summing *s*). This implies a positive cost for the principal in terms of agent's effort. At some point, *i.e.*, when $\lambda_L = \lambda_1$, this cost is more than compensated by the benefit of designing an even larger *t* and insuring also against the risk of the mid-expert serving the client.

The optimal contract thus becomes a referral contract. With further increases in λ_L, the referral contract also becomes too costly. After a certain threshold $\lambda_L = \lambda_2$, the size of the expected mismatch is so limited that it does not pay for the principal to sacrifice effort to avoid it. As a consequence, the individual contract becomes the optimal contract.

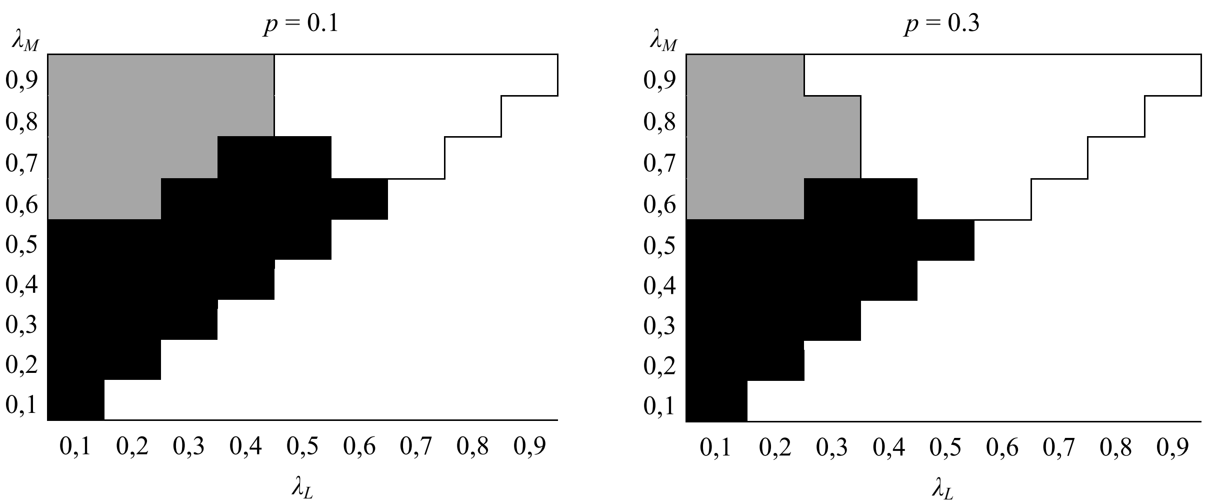
Figure 4 – Optimal contract

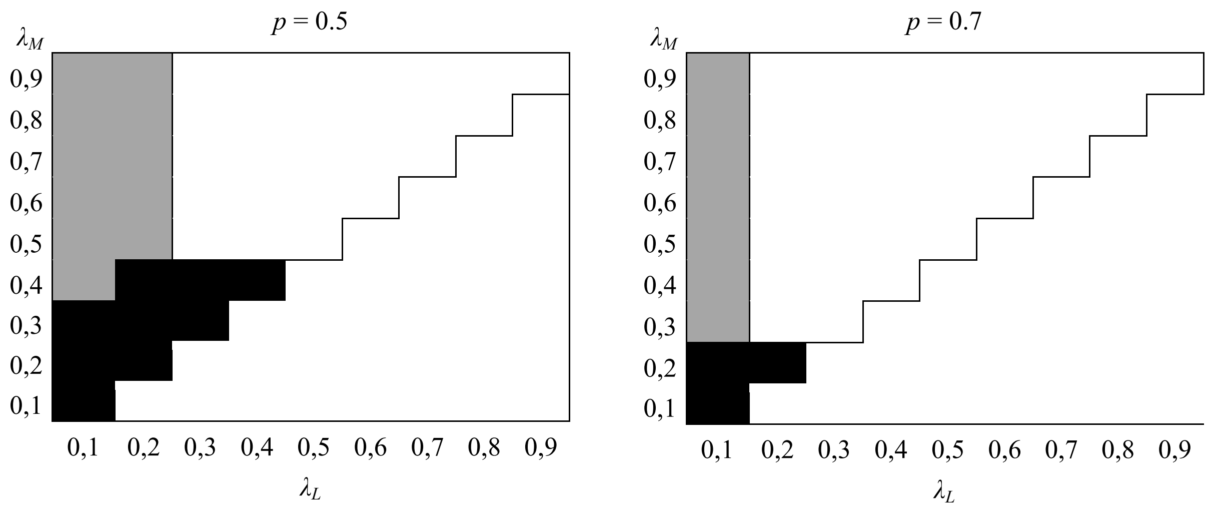


Note. The curves are computed considering the following combination of parameters: $\lambda_M = 0.7$, $p = 0.1$, $n = 5$, $\beta = 0.5$, $\theta = 1$, $\sigma = 0.5$, $x = 10$. The above panel shows that, for a non-empty set of points, each of the three contracts is the most profitable. The below panel reports changes in s and t as the frontier of the optimal contract varies.

In Figure 4 we discuss changes in optimal contract for a very limited portion of the parameter space. One may wonder what happens in the remaining portions. To address this concern we report in Figure 5 the results of several simulations where we vary all the three parameters of interest, namely λ_M, λ_L and p. Each cell in the matrices is colored in a different way depending on the optimal contract: black when the referral contract is optimal; grey when the imperfect referral contract is optimal; and white when the individual contract is optimal. Moving from one matrix to the other, we vary the value of p increasing it from 0.1 (as in Figure 4) to 0.7. As can be seen, the portion of the (λ, p) space in which one contract does better than the other varies significantly with changes in the client's competence. When p is small (top-left matrix), the region in which the individual contract is optimal is fairly limited, and is restricted to the cases in which there is no high difference in specialization among the agents, *i.e.*, both λ_M and λ_L are relatively high. In all the other regions, the expected cost of mismatching is too high and either the referral contract or the imperfect referral contract is preferable. As p increases, however, the client becomes competent at identifying the correct expert for her needs so that the expected mismatching reduces. As a consequence, the individual contract becomes more profitable even for lower values of λ_M and λ_L. In these cases it pays to incentivize effort rather than to avoid the (small) mismatching. Obviously, we find that the more we move towards the top left corner of each matrix, the more the imperfect referral contract is preferable. In these cases, in fact, the specialization of the mid-expert is so close to that of the expert that it pays for the principal to only insure against the risk that the non-expert serves.

Figure 5 – Changes in optimal contract





Note. In the black area, the Referral Contract is the most profitable; in the grey area, the Imperfect Referral Contract is the most profitable, and in the white area, the Individual Contract is the most profitable. Simulations have been run for the following combination of parameters: $n = 5$, $\beta = 0.5$, $\theta = 1$, $\sigma = 0.5$, $x = 10$.

Overall, Proposition 3 together with Figures 4 and 5 reinforce the key message derived from Section 4. In a highly specialized setting, the principal needs to take into account a trade-off between inducing more effort and avoiding mismatch when he designs the contract. A referral contract works just like a comprehensive cover insurance in which the principal eliminates all the risk that some mismatching occurs. As with all comprehensive cover insurances, however, this choice comes at a cost. In our setting, the cost is related to higher t (and consequently the lower s) that has to be paid. Obviously, whenever the risk that some mismatching occurs is low (*i.e.*, p is high), and/or its cost is limited (*i.e.*, λ_M and/or λ_L are high), it is more profitable for the principal to either resort to a partial cover insurance, *i.e.*, to design an imperfect referral contract, or to avoid insurance altogether, *i.e.*, to design an individual contract.

6. Shirking condition

In Sections 4 and 5 we solved the model focusing on symmetric equilibria with no shirking. In a referral contract, however, an agent could be tempted to shirk, that is to reduce the effort exerted at Stage 2. Suppose an agent fully shirks and exerts a zero level of effort at Stage 2, then when he is the expert he could either refuse to serve the client if another agent refers the client to him, or he could refer the client to another agent whenever he is the first agent to be matched with the client. Following a mechanism design approach, the behavior can be avoided in a simple way. First, notice that throughout the paper we have assumed that once a client has been referred to an agent, the agent has to serve her. This means that referral can occur at most once. Second, the principal can ask the agents to tell him if they serve the client when they are not the expert. If this occurs, the agent must still serve the client and his expected payoff will be the same, but then the principal will refuse to give the bonus to all the other agents (that is the contract explicitly contains a clause such that if the agent who serves the client reports to the principal that he is not the expert, then $t = 0$). Notice that the principal does not need to observe who serves the client. If all agents are selfish, a SPNE equilibrium with no-shirking at Stage 2 exists which is sustained by this credible threat: in equilibrium, all agents who are not experts report to the principal with probability one whenever someone refers the client to them. Moreover, if the agents exhibit any (positive) level of reciprocity, or inequity aversion in their preferences, reporting the truth to the principal is the unique best response, reinforcing the proposed mechanism. It is important to notice that giving the agents the possibility to send the message to the principal does not invalidate our previous results. For instance, the principal cannot use a mechanism to elicit the information about the identity of the referral in order to give a referral bonus only to this agent. If the principal asks this information to the agent who serves the client, he can extract all the rent by auctioning which agent to indicate as referral, destroying any incentive to make a referral.

Therefore, shirking can only arise when we restrict the principal from eliciting any information from the agents, and to only propose compensation schemes contingent on observable variables. Even in this case, we show that there always exists a non-empty set of points in the (λ, p) space such that a referral contract is more profitable for the principal than an individual contract. For the sake of simplicity, we analyze this case under the assumption that only two levels of λ exist. The results would also hold in the case of multiple λ.

Suppose an agent can shirk. In this case, he refuses to serve a client. If he is the first agent to be matched with the client, he will refer the client to a colleague to receive the team bonus with positive probability. To be incentive compatible, the referral contract must thus satisfy an additional "no-shirking" condition. Let e^R be the amount of effort exerted by any agent in a referral contract and assume as in Section 4 that all non-expert agents have the same ex-ante specialization λ. A referral contract satisfies the non-shirking condition if:

$$\frac{1}{n}\left(e^R\right)^{\beta} s^R - c(e^R) \geq \lambda\left(e^R\right)^{\beta} t^R \tag{5}$$

where the previous inequality follows from the fact that whenever an agent deviates and reduces his effort to a level such that when he is the expert, he is not ex-post the agent with the highest probability to succeed, then the most profitable deviation is $e = 0$. On this basis, it can be shown that:

PROPOSITION 4:

a) If $\lambda \leq [(2-\beta)/2n]^{1/2} = \underline{\lambda}$, *then the optimal referral contract identified in Proposition 2 satisfies the non-shirking condition.*

b) A non-empty set exists in the (λ, p) *space such that the optimal referral contract satisfies the non-shirking condition and is more profitable than an individual contract.*

Proof: See Appendix.

If agents can shirk, the optimal referral contract satisfies the non-shirking condition only if the fitness of the non-expert is sufficiently low. By shirking an agent in fact knows that he is not going to serve the client when he is the expert, and therefore the cost of shirking is larger the lower the probability of receiving the team bonus when the non-expert serves. Notice that, not surprisingly, the non-shirking condition is harder to satisfy in large firm: as n increases $\underline{\lambda}$ reduces.

7. Conclusion

Specialization is required in modern firms. Employers encourage their employees to develop specific and narrow specialization to be able to better satisfy clients' needs. However, this specialization has a cost: clients may not be able to identify the correct expert and they may therefore contact the wrong employee. The employer has to take into account the resulting (costly) mismatch and when this cost is severe, the principal has to specify a referral contract in order to solve the mismatching problem.

With this paper we explore the possibility that the client addresses the wrong agents and we shed new light on the cost of asymmetric information between the client and agents for the principal. Our results show the trade-off between incentivizing effort and giving agents the incentive to refer, while we also identify under which conditions a Referral Contract (RF) is more profitable than an individual Contract (IC). In our analysis we have also studied an imperfect referral contract (IRC) that only mitigates the mismatching problem by only solving the mismatches that have the highest cost for the firm.

In our model we do not consider repeated interactions among agents or between agents and clients. Once the client has applied to the firm and has been served (either successfully or otherwise) the game ends. In more complex settings one could imagine a flow of clients contacting different agents in the firm. In these cases, implicit cooperation among agents could eliminate some degree of mismatching even when an individual contract is used. For instance, a non-expert agent could decide to refer the client to the expert in the expectation that, when he happens to be the expert in some future round, this act will be reciprocated by his colleagues. However, in more complicated settings, reputation is not enough to ensure that the cooperative equilibrium is sustained. The intangible nature of a referral act may indeed be the cause of information asymmetries that in turn undermine the credibility of individual commitment to cooperate. For instance, the existence of imperfect information on the flow of clients and/or on the value of the different opportunities may reasonably reduce the possibility to credibly commit to future reciprocation. Similarly, the existence of some asymmetry among the agents in the probability that they are contacted by the client would make a cooperative solution difficult to be maintained. In all these cases, our results still hold.

Finally, we want to conclude this paper by suggesting possible interventions open to the principal to reduce the severity and cost of the mismatching problem. As our model shows, the principal can reduce the size of the mismatch by increasing client competence and *sophistication*. Natural tools to achieve this are, e.g. branding or the use of the technology such as a website with questions to determine the client's needs and find the correct expert in the firm.

References

Alchian, A. A. and H. Demsetz (1972). Production, information costs, and economic firm. *American Economic Review*, 62(5), 777-795.

Blouin, M. and J. Bourgeon (2008). Practices. *Mimeo*.

Boning, B., C. Ichniowski, and K. Shaw (2001). Opportunity Counts: Teams and the Effectiveness of Production Incentives. *Journal of Labor Economics*, 25(4), 613-650.

Corts, K. S. (2007). Teams versus Individual Accountability: Solving Multitask Problems through Job Design. *The RAND Journal of Economics*, 38(2), 467-479

Darby, M. R. and E. Karni (1973), Free Competition and the Optimal Amount of Fraud. *Journal of Law and Economics*, 16(1), 67-88

Debo L. G., L. B. Toktay and L. N. Van Wassenhove (2008) "Queuing for Expert Services."; *Management Science*, 2008, 54(8): 1497-512.

Dulleck U., R. Kerschbamer and M. Sutter (2011) "The Economics of Credence Goods: An Experiment on the Role of Liability, Verifiability, Reputation, and Competition." *American Economic Review*, 101(2): 526-55.

Epstein, A.J., J.D. Ketcham and S. Nicholson (2010) "Specialization and matching in professional services firms", *The RAND Journal of Economics*, 41(4): 811-834.

Faraj S and L Sproull (2000) "Coordinating expertise in software development teams" *Management Science*, 1554-1568

Fuchs W. and L. Garicano (2010), Matching Problems with Expertise in Firms and Markets. *Journal of the European Economic Association*, 8(2-3):354-364

Garicano, L. (2000). Hierarchies and the Firm of Knowledge in Production. *Journal of Political Economy*, 108(5), 874-904.

Garicano, L. and T. Hubbard (2009). Specialization, Firms, and Markets: The Division of Labor within and between Law Firms. *Journal of Law, Economics and Organizations,* 25(2), 339-371.

Garicano, L. and T. Santos (2004). Referrals. *The American Economic Review*, 94(3), 499-525.

Hamilton, B.H., J. A. Nickerson and H. Owan (2003). Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation. *Journal of Political Economy*, 111(3), 465-97.

Harris M. and A. Raviv (2002) "Organization Design." *Management Science*, 48(7), 852-865.

Inderst, R. and M. Ottaviani (2009). Misselling through Agents. *The American Economic Review*, 99(3), 883-908

Itoh, H. (1993). Coalitions, Incentives, and Risk Sharing. *Journal of Economic Theory*, 60, 410-427

Itoh, H. (1991). Incentives to Help in Multi-Agent Situations. *Econometrica*, 59(3), 611-636

Itoh, H. (1992). Cooperation in Hierarchical Organizations: An Incentive Perspective. *Journal of Law, Economics and Organization*, 8(2), 321-345

Itoh, H. (1993). Job Design, Delegation, and Cooperation: A Principal-Agent Analysis. *European Economic Review*, 8, 691-700

Kinshuk Jerath, Z. John Zhang (2010). Store within a Store. *Journal of Marketing Research*, 47(4), 748-763.

Levin J and S Tadelis (2005) Profit Sharing and the Role of Professional Partnerships, *Quarterly Journal of Economics*, 120(1), 131-171.

Malone T. W., R. J. Laubacher, T. Johns, (2011) The Big Idea: The Age of Hyperspecialization, *Harvard Business Review*, 89 (7/8): 56-65.

Macho-Stadler, I., J.D., Pérez-Castrillo (1993) Moral Hazard with Several Agents: The Gains from Cooperation, *International Journal of Industrial Organization*, 11(1), 73-100.

Park, I. (2005). Cheap-Talk Referrals of Differentiated Experts in Repeated Relationships. *The RAND Journal of Economics*, 36(2), pp. 391-411

Shumsky, R. A. and E. J. Pinker (2003). Gatekeepers and Referrals in Services. *Management Science*, 49(7): 839–856

Xu, F. Z. (2011). Commission Sharing among Agents, *Mimeo*.

Appendix

Proof of Lemma 1

The effort choice problem of each agent is as follows (since all agents are *ex-ante* identical we can drop the suffix i):

$$\max_e \frac{1}{n} e^{\beta} s - \frac{\theta e^2}{2}$$

subject to $s \geq 0,\ 0 \leq \beta < 2,\ \theta > 0$

where the first term in the objective function represents the expected return when the agent is the expert (probability $1/n$), and the second term is the cost of effort. The first order condition gives us the following best-response function for the agent:

$$e(s) = \left(\frac{\beta s}{n\theta}\right)^{\frac{1}{2-\beta}}. \tag{A1}$$

It follows that the agent's optimal effort increases when the bonus s increases, but it decreases with n. Given this the principal selects the optimal bonus to maximize his expected profit:

$$\max_s e(s)^{\beta}(x - s)$$

subject to (i) $e(s) = \left(\frac{\beta s}{n\theta}\right)^{\frac{1}{2-\beta}}$

(ii) $0 \leq \beta < 2,\ \theta > 0$

Substituting the agent's best-response into the objective function and deriving the associated first order condition, we find that the optimal bonus is equal to:

$$s^* = \frac{x\beta}{2} \tag{A2}$$

Substituting Eq. (A2) back into the agent's best-response and the principal's objective function we find that the optimal effort and profit are equal to:

$$e^* = \left(\frac{\beta^2 x}{2\theta n}\right)^{\frac{1}{2-\beta}} \qquad \pi^* = \left(\frac{\beta^2 x}{2\theta n}\right)^{\frac{\beta}{2-\beta}} \frac{x(2-\beta)}{2}$$

Proof of Lemma 2

In the individual contract, the principal pays a bonus s to the agent who successfully serves the client, and does not pay any team bonus, $t = 0$. In this case, the first agent who is matched with the client always serves. Given this, in Stage 2, agents solve the following effort choice problem:

$$\max_e \frac{1}{n}\left\{p + (1-p)\frac{1}{n}\right\}e^\beta s + \frac{n-1}{n}(1-p)\frac{1}{n}\lambda e^\beta s - \frac{\theta e^2}{2}$$

$$\text{subject to} \qquad s \geq 0,\ 0 \leq \beta < 2,\ \theta > 0$$

where the first term represents the expected return when the agent serves and he is the expert, the second term represents the expected return when the agent serves and he is not the expert, and the last term is the cost of effort. The first order condition for the above problem gives us the following best-response function for the agent:

$$e = \left\{\frac{\beta s}{n\theta}\left[p + (1-p)\frac{1+(n-1)\lambda}{n}\right]\right\}^{\frac{1}{2-\beta}} \qquad \text{(A3)}$$

The optimal bonus choice problem faced by the principal in Stage 1 can be thus written as follows:

$$\max_s \left[p + (1-p)\frac{1}{n}\right] e(s)^\beta (x-s) + \left[1 - p - (1-p)\frac{1}{n}\right]\lambda e(s)^\beta (x-s) \qquad \text{(A4)}$$

$$\text{subject to} \qquad \text{(i) } e(s) = \left\{\frac{\beta s}{n\theta}\left[p + (1-p)\frac{1+(n-1)\lambda}{n}\right]\right\}^{\frac{1}{2-\beta}}$$

$$\text{(ii) } 0 \leq \beta < 2,\ \theta > 0$$

where the first term in the objective function represents the expected profit when the agent who serves the client is an expert, and the second term represents the expected profit when the agent who serves the client is a non-expert. By substituting constraint (i) into the objective function, and deriving the associated first order condition, we find that the optimal bonus in the individual contract with no referral is equal to:

$$s^{I} = \frac{x\beta}{2} \tag{A5}$$

Substituting Eq. (A5) into Eqs. (A3) and (A4) we obtain the following expressions for the optimal effort and profit in the individual contract:

$$e^{I} = \left\{ \frac{\beta^2 x}{2\theta n} \left[p + (1-p)\frac{1+(n-1)\lambda}{n} \right] \right\}^{\frac{1}{2-\beta}}$$

$$\pi^{I} = \left(\frac{\beta^2 x}{2\theta n} \right)^{\frac{\beta}{2-\beta}} \left[p + (1-p)\frac{1+(n-1)\lambda}{n} \right]^{\frac{2}{2-\beta}} \frac{x(2-\beta)}{2}$$

Proof of Lemma 3

The value of the team bonus t does not affect the incentive to exert effort and therefore it is optimally fixed at the minimum level to induce referral,

$$t = \lambda s.$$

Hence, in Stage 2 the agents will solve the following maximization problem:

$$\max_{e} \frac{1}{n} e^{\beta} s - \frac{\theta e^2}{2}$$

subject to $\quad s \geq 0,\ 0 \leq \beta < 2,\ \theta > 0$

Under this contract the agent that serves the client is always the expert, and therefore what matters in the choice of the optimal effort is the probability of being the expert (*i.e.*, $1/n$). The first order condition for the above problem gives us the following best-response function for the agents:

$$e = \left\{ \frac{\beta s}{n\theta} \right\}^{\frac{1}{2-\beta}} \tag{A6}$$

The optimal bonus choice problem faced by the principal in Stage 1 can be thus written as follows:

$$\max_{s} e(s)^{\beta}[x - s - (n-1)t] \tag{A7}$$

subject to $\quad$ (i) $e(s) = \left\{ \frac{\beta s}{n\theta} \right\}^{\frac{1}{2-\beta}}$

(ii) $t = \lambda s$

(iii) $0 \le \beta < 2$, $\theta > 0$

The solution to the problem defined in Eq. (A7) gives us the optimal bonuses:

$$s^R = \frac{\beta x}{2[1+(n-1)\lambda]} \qquad t^R = \frac{\beta x \lambda}{2[1+(n-1)\lambda]} \tag{A8}$$

Substituting the two optimal bonuses reported in Eq. (A8) back into Eqs. (A6) and (A7), we obtain the optimal effort and profit in the referral contract:

$$e^R = \left\{\frac{\beta^2 x}{2n\theta[1+(n-1)\lambda]}\right\}^{\frac{1}{2-\beta}} \tag{A9}$$

$$\pi^R = \left(\frac{\beta^2 x}{2n\theta[1+(n-1)\lambda]}\right)^{\frac{\beta}{2-\beta}} \frac{x(2-\beta)}{2}$$

Proof of Proposition 2

The optimal referral contract induces more effort than the optimal individual contract if and only if $e^R > e^I$. From Lemmata 1 and 2 this condition reduces to:

$$\left\{\frac{\beta^2 x}{2\theta n[1+(n-1)\lambda]}\right\}^{\frac{1}{2-\beta}} > \left\{\frac{\beta^2 x}{2\theta n}\left[p+(1-p)\frac{1+(n-1)\lambda}{n}\right]\right\}^{\frac{1}{2-\beta}}$$

After some manipulations we obtain:

$$p < \frac{n-[1+(n-1)\lambda]^2}{n[1+(n-1)\lambda]-[1+(n-1)\lambda]^2} = p^e(\lambda) \tag{A10}$$

where $0 < p^e(\lambda) < 1$ for any $n > 1$ and:

$$\lambda < \frac{n^{\frac{1}{2}}-1}{n-1} = \lambda^e \tag{A11}$$

Conditions (A10) and (A11) together prove point a) in the proposition.

With respect to point b), the optimal referral contract is more profitable than the optimal individual contract if and only if $\pi^R > \pi^I$. From Lemmata 1 and 2 this condition reduces to:

$$\left\{\frac{\beta^2 x}{2\theta n[1+(n-1)\lambda]}\right\}^{\frac{\beta}{2-\beta}}\frac{x(2-\beta)}{2} > \left\{\frac{\beta^2 x}{2\theta n}\right\}^{\frac{\beta}{2-\beta}}\left[p+(1-p)\frac{1+(n-1)\lambda}{n}\right]^{\frac{2}{2-\beta}}\frac{x(2-\beta)}{2}$$

After some manipulations we obtain:

$$p < \frac{n-\left[1+(n-1)\lambda\right]^{\frac{2+\beta}{2}}}{n\left[1+(n-1)\lambda\right]^{\frac{\beta}{2}}-\left[1+(n-1)\lambda\right]^{\frac{2+\beta}{2}}} = p^{\pi}(\lambda) \qquad \text{(A12)}$$

where $0 < p^{\pi}(\lambda) < 1$ for any $n > 1$ and:

$$\lambda < \frac{n^{\frac{2}{2+\beta}}-1}{n-1} = \lambda^{\pi} \qquad \text{(A13)}$$

Conditions (A12) and (A13) together prove point b) in the proposition.

Solution of the model with three levels of λ.

In Stage 3 there are three possibilities to consider. If the agent is an expert, she always serves. If the agent is not an expert, there are two alternatives. If the agent is λ_L she refers if and only if:

$$\lambda_L e^{\beta} s \le \underline{e}^{\beta} t \qquad \text{(A14)}$$

where $\underline{e}$ is the effort exerted by the expert. If the agent is λ_M he refers if and only if:

$$\lambda_M e^{\beta} s \le \underline{e}^{\beta} t \qquad \text{(A15)}$$

In a symmetric equilibrium Eqs. (A14) and (A15) reduce to $t \ge \lambda_L s$ and $t \ge \lambda_M s$ respectively. Hence, for any (s,t) there are now three possibilities: a) a contract (s,t) that induces any agent who is not the expert to refer (where both Eqs. (A14) and (A15) are satisfied); b) a contract (s,t) that induces any agent to serve (where neither Eq. (A14) nor Eq. (A15) are satisfied); and c) a contract (s,t) that induces agent with λ_L to refer and agent with λ_M to serve (where only Eq. (A14) is satisfied). Once again, we find the optimal contract in each of these cases.

Individual Contract: We first look at the case in which referral never occurs, that is when $t < \lambda_L s$. This case is equivalent to the individual contract discussed in Section 4 with the exception that this time the ex-ante (i.e. before the matching) expected fit of an agent when he is not the expert is equal to: $\bar{\lambda} = \sigma\lambda_M + (1-\sigma)\lambda_L$. On this basis, agents solve the following effort choice problem in Stage 2:

$$\max_e \frac{1}{n}\left\{p + (1-p)\frac{1}{n}\right\}e^\beta s + \frac{n-1}{n}(1-p)\frac{1}{n}\bar{\lambda}e^\beta s - \frac{\theta e^2}{2}$$

$$\text{subject to} \qquad s \geq 0,\ 0 \leq \beta < 2,\ \theta > 0$$

which gives us the following best-response function:

$$e = \left\{\frac{\beta s}{n\theta}\left[p + (1-p)\frac{1+(n-1)\bar{\lambda}}{n}\right]\right\}^{\frac{1}{2-\beta}} \tag{A16}$$

The optimal bonus choice problem faced by the principal in Stage 1 can thus be written as follows:

$$\max_s \left[p + (1-p)\frac{1}{n}\right]e(s)^\beta (x-s) + \left[1 - p - (1-p)\frac{1}{n}\right]\bar{\lambda}e(s)^\beta (x-s) \tag{A17}$$

$$\text{subject to} \qquad \text{(i) } e(s) = \left\{\frac{\beta s}{n\theta}\left[p + (1-p)\frac{1+(n-1)\bar{\lambda}}{n}\right]\right\}^{\frac{1}{2-\beta}}$$

$$\text{(ii) } 0 \leq \beta < 2,\ \theta > 0$$

By solving the above problem we find that the optimal bonus in the contract with no referral is equal to:

$$s^I = \frac{x\beta}{2} \tag{A18}$$

which is the same result as in the model with two levels of specialization. By substituting Eq. (A18) into Eqs. (A16) and (A17), we obtain the following expressions for the optimal effort and profit in the individual contract:

$$e^{I}=\left\{\frac{\beta^{2}x}{2\theta n}\left[p+(1-p)\frac{1+(n-1)\overline{\lambda}}{n}\right]\right\}^{\frac{1}{2-\beta}}$$

$$\pi^{I}=\left(\frac{\beta^{2}x}{2\theta n}\right)^{\frac{\beta}{2-\beta}}\left[p+(1-p)\frac{1+(n-1)\overline{\lambda}}{n}\right]^{\frac{2}{2-\beta}}\frac{x(2-\beta)}{2}$$

REFERRAL CONTRACT: We now look at the contract that induces any agent who is not the expert to refer in Stage 3, which is when $t \geq \lambda_M s$. This case is equivalent to the referral contract when two levels of λ exist. Also in this case it is optimal for the principal to set r at the minimum level to induce referral, i.e.:

$$t=\lambda_M s. \tag{A19}$$

In Stage 2, agents solve the same problem reported in the proof of Lemma 3, which gives Eq. (A6) as the best response function. To facilitate the reader, we report the solution here below:

$$e=\left\{\frac{\beta s}{n\theta}\right\}^{\frac{1}{2-\beta}} \tag{A20}$$

The optimal bonus choice problem faced by the principal in Stage 1 can thus be written as follows:

$$\max_{s}\ e(s)^{\beta}(x-s-(n-1)t) \tag{A21}$$

subject to (i) $e(s)=\left\{\frac{\beta s}{n\theta}\right\}^{\frac{1}{2-\beta}}$

(ii) $t=\lambda_M s$

(iii) $0 \leq \beta < 2,\ \theta > 0$

The solution to the problem defined in Eq. (A21) gives us the following result for the optimal bonuses:

$$s^{R}=\frac{\beta x}{2\left[1+(n-1)\lambda_M\right]} \tag{A22}$$

$$t^R = \frac{\beta x \lambda_M}{2[1+(n-1)\lambda_M]} \tag{A23}$$

By substituting Eqs. (A22) and (A23) into Eqs. (A20) and (A21) we obtain the following expressions for the optimal effort and profit in the case with perfect referral:

$$e^R = \left\{ \frac{\beta^2 x}{2n\theta[1+(n-1)\lambda_M]} \right\}^{\frac{1}{2-\beta}}$$

$$\pi^R = \left(\frac{\beta^2 x}{2n\theta[1+(n-1)\lambda_M]} \right)^{\frac{\beta}{2-\beta}} \frac{x(2-\beta)}{2}$$

IMPERFECT REFERRAL CONTRACT: Finally, we consider the case of referral contract that induces only some of the non-expert agents to refer, which is when $\lambda_L s \le t < \lambda_M s$. Once again, since r does not affect the agent's level of effort, for the principal it is optimal to select $t = \lambda_L s$. In this case it is easy to observe that an agent refers if and only if he is a non-expert, *i.e.*, $\lambda = \lambda_L$. When the agent is a mid-expert, *i.e.*, $\lambda = \lambda_M$, he finds it convenient to serve in Stage 3. Hence, this contract falls under the category of imperfect referral contract, because although referral may occur, it is not guaranteed that it will always be the expert who serves the client.

Under this contract, agents solve the following effort choice problem in Stage 2:

$$\max_e \frac{1}{n}\left\{ p + (1-p)\left[\frac{1}{n} + \frac{n-1}{n}(1-\sigma)\right]\right\} e^\beta s + \frac{n-1}{n}(1-p)\frac{1}{n}\sigma\lambda_M e^\beta s - \frac{\theta e^2}{2}$$

$$\text{subject to} \qquad s \ge 0,\ 0 \le \beta < 2,\ \theta > 0$$

where the interpretation of the objective function is the same as in all previous cases. Notice that this time the effort of the agent is relevant also in the cases where he is the expert and is referred to by the non-expert to serve the client (with probability1 – σ), as well as when he is the mid-expert and is contacted by the client (with probability σ). The solution of this maximization problem gives us the following best-response function for the agent:

$$e = \left\{ \frac{s\beta}{n\theta} \left\{ p + (1-p) \frac{1 + (n-1)[1 - \sigma(1-\lambda_M)]}{n} \right\} \right\}^{\frac{1}{2-\beta}} \tag{A24}$$

The optimal bonus choice problem faced by the principal in Stage 1 can thus be written as follows:

$$\max_s \left\{ p + (1-p) \left[\frac{1}{n} + \frac{n-1}{n}(1-\sigma) \right] \right\} e(s)^\beta [x - s - (n-1)t] +$$
$$+ \left\{ 1 - p - (1-p) \left[\frac{1}{n} + \frac{n-1}{n}(1-\sigma) \right] \right\} \lambda_M e(s)^\beta [x - s - (n-1)t] \tag{A25}$$

subject to (i) $e = \left\{ \frac{s\beta}{n\theta} \left\{ p + (1-p) \frac{1 + (n-1)[1 - \sigma(1-\lambda_M)]}{n} \right\} \right\}^{\frac{1}{2-\beta}}$

(ii) $t = \lambda_L s$

(iii) $0 \le \beta < 2,\ \theta > 0$

Once again, the principal selects the optimal bonus s to maximize profit given the probability that the agent who serves the client is the expert (first term) and the probability that the agent who serves the client is the mid-expert (second term). The solution to the problem specified in Eq. (A25) gives the following result for the optimal bonuses:

$$s^{IR} = \frac{\beta x}{2[1 + (n-1)\lambda_L]} \tag{A26}$$

$$t^{IR} = \frac{\beta x \lambda_L}{2[1 + (n-1)\lambda_L]} \tag{A27}$$

By substituting Eqs. (A26) and (A27) into Eqs. (A24) and (A25) we obtain the following expressions for the optimal effort and profit in the case with imperfect referral:

$$e^{IR} = \left\{ \frac{\beta^2 x}{2n\theta[1 + (n-1)\lambda_L]} \left\{ p + (1-p) \frac{1 + (n-1)[1 - \sigma(1-\lambda_M)]}{n} \right\} \right\}^{\frac{1}{2-\beta}}$$

$$\pi^{IR} = \left(\frac{\beta^2 x}{2n\theta[1 + (n-1)\lambda_L]} \right)^{\frac{\beta}{2-\beta}} \left\{ p + (1-p) \frac{1 + (n-1)[1 - \sigma(1-\lambda_M)]}{n} \right\}^{\frac{2}{2-\beta}} \frac{x(2-\beta)}{2}$$

Point a) in Proposition 3 follows from the direct comparison of Eqs. (A18), (A22), (A23), (A26) and (A27). Point b) is self-explanatory.

Proof of Proposition 4

Point a) is derived by substituting Eqs. (A8) and (A9) into the non-shirking condition defined in Eq. (5). Point b) follows directly from Proposition 2 and the fact that $0 < \underline{\lambda} < 1$ for any $n > 1$ and $0 \leq \beta < 2$.