NBER WORKING PAPER SERIES

REFERRALS

Luis Garicano
Tano Santos

Working Paper 8367
http://www.nber.org/papers/w8367

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2001

We are grateful to Canice Prendergast for many discussions and insightful suggestions. We also benefited greatly from conversations with Ulf Axelson, Pierre-Andre Chiappori, Wouter Dessein, Doug Diamond Robert Gertner, Tom Hubbard, Glenn MacDonald, Kevin Murphy, Michael Raith, Rafael Repullo, Matthew Rhodes-Kropf, Antoinette Schoar, Pablo Spiller, Lars Stole, Tim Van Zandt, and David C. Webb. We also thank participants in seminars at the University of Chicago, Stanford Institute of Theoretical Economics (SITE), Berkeley, INSEAD, LSE, Columbia University, MIT, UCLA, Duke, Rochester, CEMFI, and at the Society of Economic Design in Instanbul. We are grateful to the GSB for financial support. Part of this research was conducted while the second author was enjoying the hospitality of the Finance group at the Graduate School of Business of Columbia University. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

Referrals
Luis Garicano and Tano Santos
NBER Working Paper No. 8367
July 2001
JEL No. D2, G3, L2

ABSTRACT

Specialization requires that workers deal with some valuable opportunities themselves and refer other, possibly unverifiable, opportunities to other workers. How do markets and organizations ensure the matching of opportunities with talent in the presence of informational asymmetries about their value? The cost of providing incentives for effort in this context is that they increase the risk of the agent appropriating an opportunity she should refer upstream. Thus spot markets are severely limited in their ability to support referrals, as they involve very powerful effort incentives on those opportunities kept by the referring agents. We show that partnerships, in which agents agree to share opportunities and the income from the opportunities, appear endogenously as a solution to this problem. Partnership contracts support better communication rules at the expense of biasing effort provision away from first best for all activities. The structure of the contract depends both on the frequency of communications and on the interaction between the relative skill of the agents and the direction of the referral flow.

Luis Garicano
Graduate School of Business
University of Chicago
1101 East 58th Street
Chicago, IL 60637

Tano Santos
Graduate School of Business
University of Chicago
1101 East 58th Street
Chicago, IL 60637
and NBER

I. INTRODUCTION

> "The desirability of creating organizations of a scope more limited than the market as a whole is partially determined by the characteristics of the network of information flows." Arrow (1974:37).

A crucial task of economic organization is the match of valuable opportunities with talent. However, because of the decentralized nature of information, no central authority is able to determine who is best qualified to deal with a particular problem.[1] In fact, often the only individual who knows that an opportunity exists is the one confronting it.

This paper studies how economic organization assigns opportunities to talent when the knowledge of the time and place of an opportunity is separate from the substantive knowledge required to deal with it. It shows the conditions under which a partnership, a set of agents who agree ex-ante to share the opportunities that may arise and the income from these opportunities, naturally emerges as a solution to this assignment problem.

The problem we study is of particular relevance to the organization of professional services. Consider for example a mediocre personal injury lawyer who receives a potentially highly profitable client.[2] The lawyer realizes that, were she better qualified, she could obtain a substantially higher profit from the opportunity. The problem is that she is the only one who knows this and, absent any trade, she stands to obtain a considerable, but lower, profit from this client.

Instead, she could refer the client to a better lawyer than herself. The obstacle to this transfer is that the contract generally cannot be made contingent on the actual quality of the case, since determining the value of the claim requires meeting the client, with a high risk that the client is lost to the originating lawyer. As a result, absent reputational considerations, the market may be prone to adverse selection, as the referring lawyer may try to refer bad cases as if they were good.

[1] It was Hayek (1945:519) the first who stressed that "the peculiar character of the problem of a rational economic order is determined precisely by the fact that the knowledge of the circumstances of which we must make use never exists in a concentrated or integrated form."

[2] We present systematic empirical evidence on contracts in the personal injury referrals market later in the paper.

To investigate these issues we write a simple model where clients lack the expertise to assess the value of their claim. As a consequence, clients cannot sort themselves efficiently. We capture this problem by having opportunities randomly flow to agents. Agents then must decide between passing on the opportunity or handling it themselves. Team production is endogenous, as the decision to involve another agent in the production process lies with the agent who knows about the economic opportunity.[3] Before engaging in production the problem of the client must be diagnosed in order to learn its full value. The information obtained from the diagnosis is unverifiable to outsiders, as it is based on 'soft' knowledge. Allocating optimally the opportunity requires compensating the agent who knows about its existence in a way that gives her the proper incentives when she is actually best qualified to deal with it, while giving her incentives to refer the rest. At the same time, the contract must avoid moral hazard on the part of the agent receiving the referral, who could always blame the referring agent for misreporting the value of the opportunity.

When opportunities accrue to high skill agents, referrals flow, if they do, "downstream," that is, from more skilled to less skilled agents. In this case referral incentives are perfectly aligned with efficiency, as the high skill agent prefers to keep the valuable tasks and to refer downstream the rest. As a result, simple fixed price contracts implement the first best allocations. Similarly, no incentive conflict exists when the matching problem is strictly 'horizontal,' that is, whenever agents' skills are not compatible with each others' opportunities. In this case the referring agent cannot appropriate the opportunity and fixed price transactions support the first best.

We show that this simple result does not survive when referrals flow "upstream," that is, from less to more skilled agents. Now fixed price transactions lead to adverse selection problems and market breakdowns, as any price that induces a less skilled agent to refer valuable opportunities leads her also to refer less valuable ones. Sharing contracts may be used to signal the value of the opportunity. However, since the referring agent can always choose to keep the problem herself, the necessary distortions in the output of the high skill

[3]Thus our model combines aspects of a multitasking problem (Holmstrom and Milgrom 1987, 1991) with those of team production (Holmstrom, 1982). It is multitasking problem because agents have to decide between referring the opportunity or engaging in production. It is also a team production problem because, if the referral takes place, one agent will perform the task of referral and another that of production.

agents are such that the most valuable problems may not be referred. As a result, the spot market severely under-refers compared to the first best.

When the agents can monitor each other's income even when no joint production is involved, they can improve on this allocation by establishing an ex-ante referral agreement or 'partnership.'[4] These contracts commit them to share income from their clients regardless of who ultimately handles the opportunity. In particular they allow the partnership to "tax" the agent in possession of the opportunity even when she does not refer it. As a consequence, these contracts reduce effort incentives on the opportunities that the agent retains but, at the same time, they also reduce her incentives to hold on to the opportunities ex-post, making her willing to refer them for a lower share.

Writing such a complete contract requires making income from the clients observable and verifiable, even when the same agent who draws the client deals with the opportunity and no joint production takes place. In turn, making the revenue flows observable and verifiable requires that the agents jointly bill their clients. This joint-billing feature is likely to lead the ex-ante referral agreements object of our study to take the form of legal partnerships or firms.[5]

Professional service partnerships are, in this view, joint billing arrangements that allow agents to create a nexus of ex-ante revenue and opportunity sharing contracts with each other. Such partnerships should be created by agents when some threat of appropriability of the value of the opportunities exist, i.e., when the skills of the agents are specialized but to some extent overlaping, so that agents can extract some rents by withholding opportunities that should correspond to another agent. No specific assets or non-contractible investments (unlike in the Grossman-Hart-Moore setting) need to exist. Agents form partnerships because specialization requires sharing opportunities. This sharing is facilitated by ex-ante contracting, as individuals have a smaller incentive to extract information rents by withholding opportunities from those who are better suited to take advantage of them. Ex-ante

[4]The economics literature (e.g. Farrell and Scotchmer, 1988 or Legros and Matthews, 1993) has used the word partnership to refer to ex-ante income sharing arrangements between agents engaged in team production. Our usage, which refers to agents who agree to share opportunities and the income from those opportunities, is more closely aligned with the legal term.

[5]A priori, it is possible that two agents could agree to bill their clients together through the same accountant, for example. In our view, this would be, the facto, a partnership.

contracting requires, in turn, jointly billing the clients of the agents involved in order to avoid concealment of the existing opportunities.

An additional implication of our model refers to the determinants of effort incentives faced by the agents in partnerships. In the standard setting, the cost of providing incentives is that they force the agent to bear some of the risk associated with production, which is costly because of either risk aversion or limited liability. However the evidence supporting this traditional view is "tenuous" at best (Prendergast (1999,2000)). In contrast, we suggest that the cost of providing incentives for effort is that they increase the risk of the agent missappropriating an opportunity she should refer upstream. In particular, incentives for effort provision are more powerful the lower the expected need for referrals.

No previous literature has, to our knowledge, studied referrals under asymmetric information. Garicano (2000) discusses a hierarchical referral process when agents incentives are aligned, so that asymmetric information is not a problem. Demski and Sappington (1987), Wolinsky (1993) and Taylor (1995) deal with the role of expert advisors under asymmetric information, but all these papers study the relation between the client and the expert, rather than the incentives of the expert to allocate the problem to someone more qualified.[6] We assume instead that clients are completely uninformed and arrive randomly to experts, and focus on the relationship between the different experts.

Another branch of the literature deals with the behavior of agents in partnerships. For Farrell and Scotchmer (1988) partnerships are coalitions that divide output equally. They study coalition formation games under such constraint, and obtain implications for the size and composition of such partnerships. Legros and Matthews (1993) study incentives in deterministic partnerships (teams sharing output) with general sharing rules and show that, under very general conditions, partners choose the efficient actions. Kandel and Lazaer (1992) study peer pressure in partnerships, taking the existence of the partnership as given. None of these papers deals with the problem of the allocation of opportunities among agents, which this paper considers their fundamental role. Moreover, these papers take the ex-ante

[6]Demski and Sappington study the problem of inducing an expert to acquire knowledge in situations when the easiest thing for him is to make a blind recommendation rather than go through the trouble of actually figuring out what went wrong. Wolinski studies the development of reputation by these experts in the presence of consumer search. Taylor analyzes the market solutions to the problem posed by the expert's incentive to always recommend treatment to an uninformed consumer.

contract between the parties committing them to share income as given. Unlike in these papers, partnerships appear endogenously in our set-up. Agents agree to share income and opportunities, even though they could choose not to share either.

An alternative hypothesis that aims to explain the emergence of professional service partnerships was first enunciated informally by Gilson and Mnookin (1985), who suggest that law partnerships are a risk pooling arrangement. The aim of this arrangement is to encourage ex-ante investment by lawyers in specialized areas which have uncertain future demand. They argue that firm specific capital is the glue that binds the firm together and prevents those that invested in successful practices from leaving the firm. We believe this treatment fails to take into account the substantial income variation that exists inside most law firms, as we show in the next section. Moreover, a risk sharing model makes the opposite prediction of our model for the scope of the partnership. Law firms should include, in such a model, as many unrelated areas as possible, in order to ensure the maximum amount of risk-sharing possible. In fact, and consistently with our predictions, law firms tend to include related areas among which ensuring the right referral flow is subject to potentially large incentive conflicts.

The paper is structured as follows. Section II presents some examples that motivate our modelling choices. Section III presents a model of upstream vertical referrals, which is investigated in sections IV and V. Section VI considers downstream and horizontal referrals, and shows that in this case no asymmetric information problem exists. Section VI also discusses the implications of the model for the scope of the partnership. Section VII concludes.

II. SOME EXAMPLES

In this section we discuss some of the evidence that motivates our modelling choices.

A. Clients of professional services may be unable to evaluate their problem and identify the right expert

Given the technical nature of the services that lawyers, consultants, accountants, bankers, and other professionals provide, consumers are not able to determine the right expert for their problem. For this reason, these professionals are frequently confronted with opportunities they are not best suited to solve. Knowledge of the field allows them to identify the agent who is best suited for that opportunity. In this situation they must choose between dealing

with the opportunity themselves or referring it to an appropriate agent. This referral may take place either in the market or inside the firm.

An example of market referrals is the injury claim referral market in the state of New York, which has been studied by Spurr (1988 and 1990). In this market, a lawyer who knows of a client with an injury claim may refer him to another lawyer. Both lawyers enter into a referral agreement which is sanctioned by the New York Bar Association.[7] Similarly, as we shall see below, law firms (Cotterman (1995)), consulting firms (Mudrick (1990)), and other professional service firms[8] have explicit compensation formulas that reward professionals for giving up opportunities and passing them on to other members of the firm in exchange for some share of the overall revenues.

Referrals are also present in the financial sector. For instance, a venture capital firm with little expertise in internet matters may know of an internet opportunity and need to refer to another one.[9] In the commercial banking sector, the issue is whether customers with substantial investment assets will be served by trust officers or the stock brokers employed by the bank, which offer almost identical products. Banks establish thresholds of investable

[7]These contracts are at the root of multiple legal disputes. As a result, court verdicts are also an excellent source of documentation on the nature of these contracts. A recent example of these disputes is *Florida Bar vs. Kevin Carson*, (91,550 Florida (1998)), in which the dispute is the result of an agreement whereby 'the two lawyers entered into a mutually advantageous referral relationship, whereby [one of the lawyers] Mr. Vasilaros told [the other lawyer] Mr. Carson that he would pay Mr. Carson 25% of the attorney's fee for personal injury cases that he obtained as a result of referrals made by Mr. Carson to him.' A high profile dispute is documented in 'Against O'Quinn: Ex-partner sues lawyer for $250 million' (*Houston Chronicle* Feb 15, 1999). Among the issues at the heart of the dispute, the plaintiff claims that the defendant "took cases referred by [the defendant's] associates but did not pay him from these cases as he was supposed to.'

[8]Referrals also occur among sports and other talent agents. A sport or talent agent may know of a hot rookie player or actor who would be better off in the hands of another agent with better contacts, although the rookie may ignore that. For a some discussion of how sports agents obtain a client list (through effort or referral) see 'Most black agents shy away from America's pastime' Jerome Solomon, in *Houston Chronicle*, June 27, 1999.

[9]Recently, some internet firms have been created to improve the matching between venture capitalists and firms by facilitating the referral process, notably Venture Capital Online (vcapital.com). Guides to seeking Venture Capital routinely recommend those seeking funds to explicitly ask venture capitalists they contact for a referral to other venture capitalist that may be more suitable (see e.g. http://www.venturedirectory.com/venturecapital.htm).

assets that segment their customer base into trust or securities brokerage customers, with customers above the cut-off point being assigned to trust.[10] However brokers frequently find customers who would belong to trust officers according to the internally decided cut-off point.[11] Banks must provide incentives to brokers to refer these customers to the trust department.

We emphasize that the referral practices in these examples arise because *only the referring agent,* and not the client or the professional who receives the client from him, has any information about the quality of the claim. Clearly the client lacks the expertise to assess the value of his claim and cannot self refer efficiently, whereas the receiving professional does not have access to the information of the referring expert. This informational asymmetry is at the root of the contracting problem we discuss next.[12]

B. Referrals between agents in similar product space and geographic areas take the form of sharing arrangements; those between agents in clearly differentiated markets do not involve sharing

The estimation of the value of an opportunity is often based on soft information. This inherently complicates its truthful communication. Moreover, and maybe more importantly, revealing the existence of the opportunity means, in many instances, giving it away. The

[10]On the referral problem in the banking sector see Kehrer (1998) who reports thresholds, that vary across banks, between $200,000 and $3,000,000.

[11]This occurs because, according to Kehrer (1998, page 98), in many banks "the security brokerage sales force dwarfs the trust sales force."

[12]Another important case in which referrals play a large role is among physicians. In this case self-referral is usually impossible. We do not discuss this case however in the text, as writing explicit referral contracts on the value of the output ('fee splitting') is explicitly forbidden by the AMA, 36 state laws and the Federal Medicare and Medicaid statute (Rodwin 1995:117-120). Still, referral contracts did exist in the past. Even though in 1902 the AMA resolved that fee splitting without patient knowledge would be considered a misconduct, this prohibition was not enforced at the time. In 1912, this organization explicitly made fee splitting acceptable so long as it was disclosed. However, since the foundation of the American College of Surgeons (who were the ones paying the kickback) in 1913, a powerful anti-splitting force was created. Members signed an oath to "shun unwarranted publicity, dishonest money seeking and commercialism" and "refuse utterly all secret money trades with consultants and practitioners" (Rodwin, 1995:29). Between 1914 and 1953, 22 states passed status making fee splitting illegal (see Rodwin, 1995:23-26). For a brief review of the history of financial conflicts in the medical profession, with a particular emphasis on fee splitting see Rodwin (1992). For a survey of physicians' financial incentives see Magnus (1999).

market for knowledge of 'circumstances of time and place' is in this respect similar to the market for other types of knowledge in which, as Teece (1996:107) has pointed out, 'transactions [...] must proceed under conditions of ignorance.'

As a result, contracts referring opportunities cannot be based on the opportunity itself. We observe empirically two regularities:

B.1. Referral contracts between agents with similar skills and similar geographic markets take the form of output-sharing arrangements.

Spurr (1988, 1990), studies referrals among personal injury lawyers. He finds that the referring and receiving lawyer divide up the income obtained from the claim. The contract only specifies the output shares corresponding to each lawyer, and does not bind any of the parties to devote a minimum amount of time or effort to pursuing the claims. Spurr documents[13] the existence of referral contracts involving surprisingly substantial sharing of the recovery of the claim between the referring lawyer and the one who ends up doing the work. Furthermore, there is a large cross sectional variation in the referral shares of those contracts. Table I below, which is taken from Table 4 of Spurr (1988), reports the distribution of referral shares for the contracts in Spurr's sample.

Similarly, contracts inside accounting,[14] consulting firms[15], and law firms, where agents have skills that allow them to compete for the same clients, rely on output sharing to decentralize the allocation of opportunities.[16] These firms usually reward their partners according to their performance on many dimensions, one of the most important of which is what they call business origination.[17]

[13]Spurr had access, through an order of the Federal District Court of New York, to the file retainer and closing statements. These files contain information about the terms of contract between both lawyers. In particular they contain the fees to be earned by each lawyer, the gross recoveries and the share of the recovery assigned to the (actual) litigation lawyer; whether the lawsuit was filed, settled, went to verdict, and the verdict itself and the names of the lawyers.

[14]For a discussion see Mudrick (1997).

[15]Personal communication from several partners of economics consulting firms.

[16]For references and data on law firm's compensation schemes see Cotterman (1995) as well as the survey of Compensation Systems in Private Law Firms by Altman, Weil, and Pensa (2000). Additional evidence can be found in the survey by The Commercial Lawyer, (June/July 2000 issue). For an insightful discussion of these topics see also Gilson and Mnookin (1985).

[17]As emphasized by Wilber (2000) business origination, or "rainmaking," is the most used criterion in

This is done using either subjective or objective criteria. Most law firms which link income to performance rely on subjective performance evaluation (82% according to Cotterman, 1995:24), making it hard for the outsider to quantify the extent to which 'origination' is rewarded. These firms usually rely on either peer evaluation or some management committee to determine the rewards.[18]

Table I

Referral shares in the market for injury claims

Referral share	$\frac{1}{6}$	$\frac{1}{3}$	.4	$\frac{1}{2}$	.6	$\frac{2}{3}$
Share of contracts (%)	.71	27.56	10.25	52.6	4.59	.71

Source: Spurr (1988)

Other law and accounting firms explicitly rely on objective distribution systems that reward referrals and business origination. The most extended system, known as Hale & Dorr system,[19] distributes firm income to each partner according to their contribution to the firm's profits. This system credit partners for *work done* (60%), *business origination* (30%), and *profit credit* (10%) (Cotterman 1995). Other systems rely directly on business

compensation decision making, above fees collected. Still, not all firms rely on performance measures to compensate partners. Some (the minority, according to Cotterman, 1995:29) rely on the lock-step method, in which only seniority enters in the calculation of the rewards of the partners. Of the 386 US law firms covered in the survey by Altman, Weil and Pensa only 9% indicated use of the lock-step method (for a brief overview of this survey see Wilber (2000)). Also, according to The Commercial Lawyer, (June/July 2000 issue) of the 20 top law firms in the US, only four use the pure lock-step method: Cleary, Gotlieb, Steen & Hamilton, Cravath, Swaine & Moore, Davis Polk & Wardwell, and Wachtell, Lipton, Rosen & Katz. The remaining ones used performance related methods similar to the ones we discuss in the text, that go from the mild modified lock-step method of Latham & Watkins (85% lockstep and the rest is performance) to pure merit based methods. Incentives to share business and provide effort in lock-step law firms are provided by the law firm's "culture," according to personal interviews with a Cleary partner. For a brief discussion of the advantages and disadvantages of each system see Altonji (2000).

[18]Some of these evaluation systems are a *ballot*, a secret peer review system, *point accumulations,* a score card system with several attributes where points are awarded by managing partners, and *direct assignment,* the most subjective of the evaluation systems where the whole compensation is decided by managing partners.

[19]After the originator of this system Reginald H. Smith managing partner of Hale & Dorr of Boston in the 1940's.

origination, and award a share of the income to the partner who can claim to have brought the client to the firm (Mudrick, 1990).

Incentive contracts in the banking sector are also in the form of output shares. In the context of the example above, brokers receive 10% of the first year's year trust fees for the referral of a customer (see Kehrer (1998), page 98).

B.2 Referral contracts between agents with entirely different skills or geographic markets do not rely on output-sharing arrangements.

Law and other professional service firms form referral networks that allow one member of the network to refer to another clients whose problems fall outside their area of expertise or jurisdiction. A consulting firm that advises law firms has identified 300 such national and international law-firm networks.[20] For example, a client may need advice on a merger with a Brazilian company, or may need to register a trademark in different countries.[21] In this case, the law firm who typically handles the clients legal needs may have no expertise in these areas and may choose to refer to another law firm with experience in Brazil or in international trademark registration.

The referral contracts inside these networks are strikingly different than the ones discussed previously. According to an in-depth report in the New York Times (June 8, 2001), the contractual structure is very simple. Once a client is referred, each firm separately charges for its services. Unlike in the arrangements discussed above, "fees are normally not shared among the firms, and law firms within networks do not charge for referrals."

C. Professional service firms care about balancing incentives for referrals and for effort provision

The professional service firm's existence depends to a large extent on the proper and often delicate equilibrium between incentives for referrals and the provision of effort. Too low shares to originating partners may lead to underreferrals, or in the parlance of law firms, "hoarding of cases;" too high shares may lead to overreferrals and too little effort by receiving partners. Thus the professional service firm attempts to solve several incentive

[20]The consulting firm is Altman Weil Inc. as represented by principal Charles A. Maddock, quoted in "Making a Network of Lawyers" by Jonathan D. Glater in *The New York Times* of June 8, 2001.

[21]Both of these are actual examples from the previously quoted New York Times article.

problems simultaneously. It tries to encourage the efficient allocation of clients by properly crediting rainmaking activities, without severely distorting the incentives of those doing the work.[22]

The lack of proper referral incentives may lead to severe conflicts within the professional service firm. An illustrative example is the fall of Watson, Leavenworth, Kelton & Taggart, a premier Park Avenue law firm considered one of the giants of patent and trademark law (see Weingarten (1981)). The firm disappeared when "the client-share system encouraged Watson, Leavenworth, Kelton & Taggart, lawyers to guard their clients affiliation against intrusion by others, creating an atmosphere of competition among partners. Some partners suspected others of hoarding cases..." The law firm eventually died when one of its star rainmakers, decided to leave the law firm and "would likely take substantial accounts with him, including the Nestle Co. Inc., one of Watson, Leavenworth, Kelton & Taggart's biggest clients."

Underreferrals are also a concern in the banking example introduced above. As Kehrer (1998) asserts, "stock brokers are prone to customer hoarding; having found a customer with significant assets to invest the bank's broker is naturally loathe to give the customer up [to the trust department]." The issue is in fact key to the internal organization of the bank: "One response is organizational. Several banks have reorganized their business units to bring retail investment services and trust under the same uniform reporting structure." (Kehrer (1998), page 95).

III. A MODEL OF VERTICAL REFERRALS

We start by discussing the optimal referral rule of a relatively less skilled agent who obtains clients for whom she is potentially underqualified. We call this an upstream vertical referral. We leave for section VI the case where clients flow from the more skilled agent or when the agents skills are so different that no substantial risk that the referring agent may appropriate an opportunity exists. We call the latter case a horizontal referral.[23]

[22]Importantly, the design of a proper compensation for referrals encourages specialization, by providing insurance to the highly specialized lawyer who own less of the client and relies more on others for referrals. On this point see Trotter (1997), pages 50-1.

[23]We use the terms by analogy to the vertical and horizontal differentiation terms in Industrial Orga-

A. Arrival of opportunities, comparative advantage, and referrals

Consider an uninformed client who walks into the office of an agent (such as a lawyer) whose skill dealing with the opportunity v that the client represents is low. We call such an agent the *low skill agent*. The opportunity v is of uncertain value, $v \in \{v_0, v_1\}$, where v_0 is drawn with probability π and v_1 with probability $1 - \pi$.[24] The agent must first diagnose the opportunity and then apply effort e to realize output. Alternatively, after diagnosis, she may refer the problem to an agent who is more skilled at dealing with that opportunity, the *high skill agent*, who will then apply effort to it. We assume that there is a measure 1 of each of both low and high skill agents.

Let the skill of the low skill agent be $\theta_l = \eta$ with $\eta \in (0, 1)$ and the skill of the high skill agent be $\theta_h = 1$. We further assume that there exists a complementarity between effort, skill and the value of opportunities, so that the marginal value of effort and talent is higher in higher value opportunities. Furthermore output is non stochastic and is given by:

$$y_i = \theta_i e v.$$

The effort cost is given by $\psi(e)$ with $\psi' > 0$ and $\psi'' > 0$. Furthermore, $\psi''' \geq 0$, an assumption needed to guarantee that the ex-ante contract we introduce below is a concave program.[25] We assume that agents' utility is given by

$$u_i(e, v) = I_i(e, v) - \psi(e),$$

where $I_i(e, v)$ is the income of an agent of type i who draws opportunity v and chooses effort e.

---

nization, where horizontal differentiation refers to characteristics that are differently preferred by different consumers, such as products placed in a Hotelling (1929) line and vertical differentiation refers to 'quality' differences among products, where a clear ranking of better and worse products exists, such as in Rosen's (1972) hedonic model.

[24] A previous version of this paper, available from the authors, analyzed the case of a continuum of opportunities and obtained substantially similar results to the ones presented here. We thank P.A. Chiappori for suggesting that we focus on the discrete case, which greatly simplifies the exposition of the economics of the problem.

[25] $\psi''' \geq 0$ is not needed to prove any of the results with the sole exception of proposition 5. This is similar to the assumptions made by Laffont and Tirole (1993) in a different but related problem (see proposition 1.1, page 59.)

There are two situations under which the agent can find herself without an opportunity: A low skill agent who has referred the drawn opportunity to a high skill agent, or a high skill agent who has not received a referral from the low skill agent. Agents who do not tackle an opportunity gain a reservation utility which we denote by $\overline{u}_h$ for the high skill agent and $\overline{u}_l$ for the low skill agent. We are interested in economies where, absent any informational concerns, comparative advantage leads to specialization. In such situations, the first best prescribes the referral of the high value opportunity, v_1, and the non referral of v_0, that is economies for which,

$$\max_e \{\eta e v_0 - \psi(e)\} + \overline{u}_h > \max_e \{e v_0 - \psi(e)\} + \overline{u}_l \tag{1}$$

$$\max_e \{\eta e v_1 - \psi(e)\} + \overline{u}_h < \max_e \{e v_1 - \psi(e)\} + \overline{u}_l. \tag{2}$$

We assume then that (1) and (2) hold throughout. (1) says that in the event v_0 is drawn, the outside value upstream, $\overline{u}_h$, should be protected and the low value opportunity should be assigned to the low skill agent. In contrast, (2), says that if v_1 is drawn then resources upstream should be engaged in production and the opportunity assigned to the high skill agent at the expense of her outside value.

We also assume that the low skill agent prefers to tackle the low value opportunity rather than rejecting it and obtaining $\overline{u}_l$, that is,

$$\overline{u}_l \leq \max_e \{\eta e v_0 - \psi(e)\}. \tag{3}$$

Assumption (3) also guarantees that the low skill agent prefers to draw tasks rather than just enjoy her outside opportunity and, as a consequence, opportunities flow into the economy.

Under these conditions the first best allocation of effort is given by:

$$\psi'(e_l^{fb}) = \eta v_0 \quad \text{and} \quad \psi'(e_h^{fb}) = v_1, \tag{4}$$

where the subscript denotes the agent exerting the corresponding effort level, and the first best level of output is

$$y_0^{fb} = e_l^{fb} \eta v_0 \quad \text{and} \quad y_1^{fb} = e_h^{fb} v_1. \tag{5}$$

Finally, the welfare level associated with the first best is:

$$\overline{W}^{fb} = \pi \left[e_l^{fb} \eta v_0 - \psi(e_l^{fb}) + \overline{u}_h \right] + (1 - \pi) \left[e_h^{fb} v_1 - \psi(e_h^{fb}) + \overline{u}_l \right].$$

The role of the referral then is to match a higher value problem with a more skilled agent. The high skill agent not only adds more value to the problem given her talent, but also works harder, since the marginal value of her effort is higher.

B. Verifiability of opportunities and information structure

The production technology above captures the situation described in the first subsection of section II: an agent receives a client who may (or may not) have to be referred to another agent. As illustrated in section II.B, however, contracts often cannot be written conditionally on the value of the opportunity nor on the effort provided by each agent. For the receiving agent to verify the value of the client would require evaluating the client. But, in this case, the referring agent agent runs the risk of being bypassed and losing the client. This is another instance of what Arrow (1971:151) termed the *fundamental paradox of information:* 'its value for the purchaser is not known until he has the information, but then he has in effect acquired it without cost.'

Given this limitation on the information available, fixed price contracts cannot motivate the low skill agent to transfer the best opportunities and keep the worst. The argument is standard (Akerlof (1970)). Any price that is sufficient to encourage a low skill agent to refer a high value opportunity leads her to the referral of any lower value opportunity as well.

Proposition 1. No fixed price contract exists which result in the low skill agent keeping the low value problems and passing on the high value problems.

The inefficiency caused by the low skill agent referring the wrong opportunity to the high skill agent produces two related types of effects that we aim to capture in our model. First, there is the direct effect of the misallocation when high skill workers are matched with lower value opportunities. Second, there is also the economic loss produced by the waste of the information obtained from the draw, substituted by the strategic misrepresentation of the value of this opportunity produced by the referring agent. We capture this effect in our set-up by assuming that the effort decision of the high skill agent is based on the information conveyed by the low skill agent.[26]

[26]Our assumption that all the information about the value of the problem is provided by the referring agent may seem extreme. It can be relaxed by assuming that both the referring agent and the agent accepting the

Output-based contracts assign each agent i a share $s_i[y]$ of the prescribed output y. Such contracts must solve two incentive conflicts. First, contracts must ensure that low skill agents with a low value opportunity find it in their interest to keep such opportunity, rather than trying to refer it to the high skill agents (adverse selection problem). Formally, let $u_l(v_0|v_0)$ be the utility obtained by the low skill agent who draws and truthfully signals v_0. Similarly, let $u_l(v_0|v_1)$ the utility of a low skill agent who having drawn v_0 behaves as if v_1 had been drawn, trying to refer it to the high skill agent. Then, the adverse selection' constraint is:

$$u_l(v_0|v_0) \geq u_l(v_0|v_1). \tag{6}$$

Second, high skill agents may choose to supply too little effort and blame the referral (a moral hazard problem). Incentive compatibility requires that the agent who receives the referral have the right incentive to produce the prescribed effort, rather than allocate too little effort to the opportunity, reduce the output, and claim that the quality of the client was misreported by the referring agent. Formally, calling $u_h(v_1|v_0)$ the utility of a high skill agent who claims he received a low value referral, when it was high, and $u_h(v_1|v_1)$ the utility of a high skill agent who acknowledges the high value problem was received, the 'moral hazard' constraint is:

$$u_h(v_1|v_1) \geq u_h(v_1|v_0). \tag{7}$$

We study how these incentive conflicts affect referrals under two different institutional arrangements that are reflected in different timings of the contract. First, as illustrated in section II, an agent may engage in trading opportunities in an ad-hoc manner, writing a spot contract only after observing an opportunity she is not qualified to handle. Alternatively, when agents jointly bill clients, they convert the income flow in observable and verifiable even when no communication that an opportunity was generated takes place. As a result, they can form a partnership, and write a contract ex-ante that commits them to share revenues and to refer opportunities to each other regardless of who deals with the opportunity.

referral may spend resources diagnosing the problem and observing noisy signals on its value. As long as the signal of the referring agent has some informational content its communication will bias the effort choice of the high skill agent, who will combine that signal together with hers to form a posterior on the value of the opportunity at hand.

We follow most of the literature on static (one period) asymmetric information problems in making two assumptions throughout the paper. First, ex-post efficiency rules out the possibility of budget breaking in the productive relationship.[27] For this reason the share of output that the contract awards the high skill agent is $s_h[y] = y - s_l[y]$. Second, the parties can commit to the contracts they sign so that no renegotiation takes place ex-post.[28] Moreover, we restrict ourselves to deterministic mechanisms and to organizational forms that solely combine one low skill agent with a high skill agent.[29]

Throughout the whole analysis, the decision to contract, and the informational asymmetries it involves, is endogenous, and depends on the gains from specialization. The parties may always choose not to communicate and solve the opportunities in autarchy.

IV. SHARING, SIGNALING, AND REFERRALS IN THE SPOT MARKET

In this case, the two parties determine the terms of the exchange once the referring party has observed the value of the opportunity at hand and the convenience of referring it. We start the analysis by studying the specific form that the relevant incentive compatibility constraints (inequalities (6) and (7)) take in this case. Then we provide a full characterization of the signaling equilibrium program and solution, as well as some numerical examples.

A. Incentive compatibility constraints and the spot signaling program

A.1 The adverse selection problem of the low skill agent

We start by studying the adverse selection constraint (6) of the low skill agent. Assume that the high skill agent accepts a referral by a low skill agent. Let $y_h(v_i|v_j)$ be the output produced by a high skill agent who believes she has received an opportunity of quality v_j

[27] This is a standard assumption. See e.g. Holmstrom (1982) and Legros and Matthews (1993).

[28] The assumption of no renegotiation is reasonable in this context. First, the existing institutions, such as law firms, have an incentive to enforce contracts to facilitate future referrals. Second, agents have an incentive to develop a reputation for not renegotiating referral contracts in the hope of maintaining the credibility of future referral transactions.

[29] An alternative organizational form would have a low skill agent fully specialized in diagnosis and in charge of distributing tasks to either another low skill agent or the high skill agent. This would indeed solve many of the asymmetric information problems here presented but at the potentially high cost of having one of the agents not producing at all. Thus an upper bound on the distortions we study is the price of an extra worker.

when the true quality of the referred opportunity is v_i, where $v_i, v_j \in \{v_0, v_1\}$. The contract awards the low kill agent a share $s_l[y_h(v_i|v_j)]$ of the output produced. Constraint (6) requires that low skill agents whose opportunity is v_0 keep it, rather than refer it and pretend it to be opportunity v_1. In the first case the low skill agent performs the first best level of effort, whereas in the second case she gets her reservation utility plus the share of the output that corresponds to her under the deviation. That is,

$$y_0^{fb} - \psi\left(\frac{y_0^{fb}}{\eta v_0}\right) \geq \overline{u}_l + s_l[y_h(v_0|v_1)], \tag{8}$$

where we have substituted the first best effort $e_0^{fb} = \frac{y_0^{fb}}{\eta v_0}$. It remains to determine the output produced by the high skill agent under a deviation of the low skill agent, $y_h(v_0|v_1)$. The effort exerted by the high skill agent is as if the low skill agent had referred the high value task, $\frac{y_h(v_1|v_1)}{v_1}$. Thus $y_h(v_0|v_1)$ is:

$$y_h(v_0|v_1) = \left(\frac{y_h(v_1|v_1)}{v_1}\right) v_0 = y_h(v_1|v_1)\left(\frac{v_0}{v_1}\right). \tag{9}$$

Equation (8) is the incentive compatibility constraint associated with the adverse selection problem faced by the low skill agent whose draw is v_1.

What prevents setting $s_l[y(v_0|v_1)]$ low enough to discourage imitation from low skill agents with low value opportunities? The answer lies in that our problem is also one of moral hazard. From the moment the opportunity is referred a new informational problem is triggered, namely, the inability of the referring party to contract on the effort supplied by the high skill agent. If no rents are left to the low skill agent who cheats, the high skill agent may have too large an incentive to reduce output and pretend she was cheated.

A.2 The moral hazard problem of the high skill agent

We study now the moral hazard constraint of the high skill agent (7). Suppose that the low skill agent does indeed refer the high value opportunity v_1. The prescribed effort is given by $\frac{y_h(v_1|v_1)}{v_1}$. As already mentioned, if instead the low skill agent had referred task v_0, the realized output would have been $y_h(v_0|v_1)$. Thus for the high skill agent to conceal the value of the client, she must make the output look as if the low skill agent had referred v_0 rather than v_1 and she had performed the prescribed effort $\frac{y_h(v_1|v_1)}{v_1}$. The effort the high

skill agent would do to obtain output $y_h(v_0|v_1)$, given that the referred task is indeed v_1, is $\frac{y_h(v_1|v_1)}{v_1}\frac{v_0}{v_1}$, much less than the prescribed effort.[30] The incentive compatibility constraint of the high skill agent (7) is then:

$$s_h[y_h(v_1|v_1)] - \psi\left(\frac{y_h(v_1|v_1)}{v_1}\right) \geq s_h[y_h(v_0|v_1)] - \psi\left(\frac{y_h(v_1|v_1)}{v_1}\left(\frac{v_0}{v_1}\right)\right), \tag{10}$$

where the left hand side of (10) is the utility under no deviation $u_h(v_1|v_1)$ and the right hand side is the utility obtained by a high skill agent who unilaterally deviates $u_h(v_0|v_1)$ by pretending the opportunity is low v_0 rather than high, v_1.

Notice that any other effort deviation by the high skill agent produces an output level outside the set

$$\mathcal{Y} = \left\{y_h(v_1|v_1), y_h(v_1|v_1)\left(\frac{v_0}{v_1}\right)\right\},$$

and, hence, unambiguously identifies the high skill agent as the shirker. As Legros and Matthews (1993) have shown, with unlimited liability these deviations can be prevented by sufficiently penalizing the shirker. For this reason, and without any loss of generality, we ignore output levels outside $\mathcal{Y}$ throughout. Notice that the low skill agent may be deviating too, but she can only deviate by sending v_0.

Finally the contract has to be such that if a referral takes place the high skill agent obtains at least as much utility as his outside value, $\overline{u}_h$, that is:

$$s_h[y_h(v_1|v_1)] - \psi\left(\frac{y_h(v_1|v_1)}{v_1}\right) \geq \overline{u}_h.$$

A.3 The program

We are now ready to write the program that determines the spot contract. A low skill agent who draws an opportunity v_1 faces the choice between referring the opportunity or keeping it and tackling it himself. Referring it requires offering an incentive compatible contract to the high skill agent such that the high skill agent agrees to receive it in exchange for an output share. On the other hand keeping it implies operating in autarchy, in which case no incentive compatibility constraint is of concern. The problem of the low skill agent whose draw is v_1 is then:

[30] This follows immediately from the fact that $y = ev$ and equation (9).

Program $\mathcal{P}_m$

$$\max_{\{\text{refer,keep}\}} \left\{ \max_{\{s_l[\cdot], s_h[\cdot], y_h(v_1|v_1)\}} \{s_l[y_h(v_1|v_1)] + \overline{u}_l\}, \max_y \left\{ y - \psi\left(\frac{y}{\eta v_1}\right)\right\}\right\} \tag{11}$$

subject to

$$y_0^{fb} - \psi\left(\frac{y_0^{fb}}{\eta v_0}\right) \geq \overline{u}_l + s_l\left[y_h(v_1|v_1)\frac{v_0}{v_1}\right] \tag{12}$$

$$s_h[y_h(v_1|v_1)] - \psi\left(\frac{y_h(v_1|v_1)}{v_1}\right) \geq s_h\left[y_h(v_1|v_1)\frac{v_0}{v_1}\right] - \psi\left(\frac{y_h(v_1|v_1)}{v_1}\left(\frac{v_0}{v_1}\right)\right) \tag{13}$$

$$s_h[y_h(v_1|v_1)] - \psi\left(\frac{y_h(v_1|v_1)}{v_1}\right) \geq \overline{u}_h \tag{14}$$

$$s_l[y] + s_h[y] = y \quad \text{for} \quad y \in \mathcal{Y} \tag{15}$$

where we have substituted $y_h(v_0|v_1)$ by it's value as given by equation (9).

Program $\mathcal{P}_m$ then selects the output $y_h(v_1|v_1)$ and, indirectly $y_h(v_0|v_1)$, and the share function $s_l[\cdot]$ subject to the incentive compatibility constraints. Next we show that this program can be considerably simplified to yield a more tractable one.

First notice that because the measure of low skill agents with the high value task is $1-\pi$ and there is a unit measure of high skill agents, competition among the latter lowers the utility they receive under the signaling contract to their outside value $\overline{u}_h$, that is, expression (14) is met with equality and, as a consequence

$$s_h\left[y_h(v_1|v_1)\right] = \overline{u}_h + \psi\left(\frac{y_h(v_1|v_1)}{v_1}\right). \tag{16}$$

Substituting (16) into (13) we obtain

$$\overline{u}_h \geq s_h\left[y_h(v_1|v_1)\frac{v_0}{v_1}\right] - \psi\left(\frac{y_h(v_1|v_1)}{v_1}\left(\frac{v_0}{v_1}\right)\right). \tag{17}$$

We have reduced the constraints of program $\mathcal{P}_m$ to (12), (17), and (15). Clearly any incentive compatible allocation can be supported with an out of equilibrium share $s_h\left[y_h(v_0|v_1)\right]$ such that (17) is met with equality, as it can always be increased without violating (12). Then:

$$s_h\left[y_h(v_1|v_1)\frac{v_0}{v_1}\right] = \overline{u}_h + \psi\left(\frac{y_h(v_1|v_1)}{v_1}\left(\frac{v_0}{v_1}\right)\right). \tag{18}$$

Finally substituting (18) into (12) we obtain the following (single) constraint:

$$y_0^{fb} - \psi\left(\frac{y_0^{fb}}{\eta v_0}\right) \geq y_h(v_1|v_1)\left(\frac{v_0}{v_1}\right) - \psi\left(\frac{y_h(v_1|v_1)}{v_1}\left(\frac{v_0}{v_1}\right)\right) - (\overline{u}_h - \overline{u}_l)\,, \tag{19}$$

where we have substituted $y_h(v_0|v_1)$ by it's value as given by equation (9).

Program $\mathcal{P}_m$ can be then written more compactly as:

Program $\mathcal{P}'_m$

$$\max_{\{\text{refer,keep}\}}\left\{\max_{y_h(v_1|v_1)}\left\{y_h(v_1|v_1) - \psi\left(\frac{y_h(v_1|v_1)}{v_1}\right) - (\overline{u}_h - \overline{u}_l)\right\}, \max_y\left\{y - \psi\left(\frac{y}{\eta v_1}\right)\right\}\right\}$$

subject to

$$y_0^{fb} - \psi\left(\frac{y_0^{fb}}{\eta v_0}\right) \geq \Theta_h\left[y_h(v_1|v_1)\right], \tag{20}$$

where

$$\Theta_h\left[y_h(v_1|v_1)\right] = y_h(v_1|v_1)\left(\frac{v_0}{v_1}\right) - \psi\left(\frac{y_h(v_1|v_1)}{v_1}\left(\frac{v_0}{v_1}\right)\right) - (\overline{u}_h - \overline{u}_l)\,. \tag{21}$$

The function $\Theta_h\left[y_h(v_1|v_1)\right]$ is the net surplus that the high skill agent generates and the low skill agent appropriates when the latter refers the low value opportunity v_0 and the former applies effort as if the high value opportunity v_1 were the one referred. Constraint (20) simply says that the utility that the low skill agent obtains when he keeps v_0 is higher than the net surplus he could appropriate from the high skill agent when he refers it.

Finally, we introduce a piece of notation that will be useful below. Let $\overline{W}^m$ be the welfare level associated with the market allocation. Recall that the market leaves the high skill agent with an utility equal to $\overline{u}_h$. The utility of the low skill agent depends on whether v_0 or v_1 is drawn and whether the latter is referred or not. That is:

$$\overline{W}^m = \overline{u}_h + \pi\left(y_0^{fb} - \psi\left(\frac{y_0^{fb}}{\eta v_0}\right)\right) + (1-\pi)\max_{\{\text{refer,keep}\}}\left\{u_l^r(v_1), u_l^{nr}(v_1)\right\}, \tag{22}$$

where $u_l^r(v_1)$ is the equilibrium utility of the low skill agent under the referral and $u_l^{nr}(v_1)$ is the equilibrium utility of the low skill agent if the referral of v_1 does not take place.

B. Characterization of the spot signaling contract

Let y_0^m and y_1^m the output associated with tasks v_0 and v_1 respectively under the signaling equilibrium. As it can be seen in program $\mathcal{P}'_m$, description of the signaling equilibrium is fully achieved by reporting the output associated with each of the opportunities, $[y_0^m, y_1^m]$.

Proposition 2. The signaling equilibrium is of one of two types.

1. The first best can be implemented, that is, $[y_0^m, y_1^m] = \left[y_0^{fb}, y_1^{fb}\right]$
2. If the first best cannot be implemented then there are two signaling equilibrium, $[y_0^{m_a}, y_1^{m_a}]$ and $[y_0^{m_b}, y_1^{m_b}]$, such that

$$y_0^{m_a} = y_0^{fb} \quad \text{and} \quad y_1^{m_a} < y_1^{fb} \tag{23}$$

$$y_0^{m_b} = y_0^{fb} \quad \text{and} \quad y_1^{m_b} > y_1^{fb}. \tag{24}$$

Furthermore $[y_0^{m_a}, y_1^{m_a}]$ is Pareto superior to $[y_0^{m_b}, y_1^{m_b}]$

The need for truthful communication may involve distortions away from the first best. The key to this distortion is that it is not possible to simply reduce the utility of the low skill agent who refers v_0 by giving her a low share of output without distorting total output produced. Such arrangement would make it too attractive for the high skill agent to claim that she was misled. Only a reduction of the output of the high skill agent below what is optimal makes it sufficiently unattractive for the low skill worker to refer a low quality client without making it attractive for the high skill agent to claim that the referring agent is cheating.

Proposition 2 also shows that there are two possible signaling equilibria with distortions but they are Pareto ranked. Intuitively, a high quality opportunity may be signaled both by distorting effort by the high skill agent upwards or downwards. Distorting it downwards is preferred, since it achieves the same separation at lower cost. For this reason we concentrate in the signaling equilibrium s_a, which we henceforth denote $[y_0^m, y_1^m]$. As the next proposition shows the distortion is severe enough as to completely destroy the complementarity between effort and the quality of the task. That is, if the first best cannot be implemented, the higher the quality of the task referred, the lower the effort of the high skill agent.

Proposition 3. Assume that the first best cannot be implemented under the signaling contract. Then the effort exerted by the high skill agent on the task referred is a decreasing function of the quality of the task.

In summary all effort distortions occur in the set of referred opportunities, which are the ones subject to the signaling costs. Low skill agents maximize utility by their choice

of effort on the opportunities they keep, but this effort will only be optimal if the opportunity is v_0. If the opportunity is v_1 then the effort supplied by the low skill agent is $\widetilde{e} = argmax\{e\eta v - \psi(e)\}$, which is clearly less than e_1^{fb}. As a consequence *optimal effort provision occurs where it is the least valuable*, namely, in the opportunities that the first best assigns to the low skill agents. In contrast high skill agents perform suboptimal effort gearing output away from first best. The next result then follows immediately.

Corollary 4. (Under-referrals) The referral set supported by the spot market is contained in the first best referral set.

The corollary implies that when opportunities are not verifiable, the spot markets equilibrium is associated with underreferrals. Moreover, as the next example illustrates, the distortion in the effort of the high skill agent may be such that the low skill agent keeps not only the low value opportunities but also the most valuable ones.

We illustrate these results below with an example of the characterization of the spot market equilibrium. Before moving on to that, it is important to note that the analysis made remains unchanged if, in order to account for possible bounded rationality of the agents, we restrict our attention to linear sharing rules which do not depend on whether the output is the equilibrium output or out of equilibrium. In the appendix we show that a separating signaling equilibrium can also be obtained when we restrict the contracting space to linear shares whose level is independent of the output realized. Of course, efficiency losses are higher in that case.

Example. Assume that $\psi(e) = \frac{e^2}{2}$. Also assume that $v_0 = 1.4$, $\overline{u}_h = 2$ and $\overline{u}_l = .85$. Figure I reports the possible allocations in the space (η, v_1). The curve $\hat{v}^{fb}(\eta, \overline{u}_h, \overline{u}_l)$ denotes the *first best referral set:* it is the frontier of high value opportunities that should be referred to the high skill agent, that is, it is the set of v_1's that, given our parametric choices, meet assumption (2). The purpose of this example is to compare the referral set supported by the sharing contracts with the first best referral set.

There are three possible regions, a, b, and c. Referrals can only be supported in regions a and b, though in region b the first best cannot be implemented. Finally in region c the signaling equilibrium does not support any referrals.

In region a the first best can be implemented because the net surplus that the low skill agent could appropriate by misleading the high skill agent about the value of the opportunity is very low, as the task v_1 itself is not very valuable. In other words is smaller than $y_0^{fb} - \psi\left(\frac{y_0^{fb}}{\eta v_0}\right)$ as seen in constraint (20).

Holding $\eta \leq .23$ constant, and as v_1 increases, the incentives for the low skill agent to mislead the high skill agent about the value of the opportunity, as given by $\Theta\left[y_h(v_1|v_1)\right]$, increase as well. In region b that incentive is strong enough that obtaining truth telling requires distorting the output downwards. In region c the downward distortion of output is large enough as to discourage the referral completely. Recall that, as we show in proposition 3, effort is decreasing in v_1 in region b as the distortions are strong enough to undo the complementarity.

In contrast when $\eta \in (.23, .635)$ only intermediate quality problems are referred. The reason is that for v_1 sufficiently close to $\hat{v}^{fb}(\eta, \overline{u}_h, \overline{u}_l)$ the distortions are high enough as to discourage the referral, as now the low skill agent is endowed with a higher skill. As v_1 increases the referral once again becomes attractive until the effort of the high skill agent becomes so distorted as to prevent the referral once again.

Finally when $\eta \geq .635$ distortions required to sustain communication have completely erased the comparative advantage of the high skill agent.

A striking feature of this outcome is that referrals can never supported precisely when they are most valuable, that is, when v_1 is very high relative to v_0. Moreover when referrals do occur, effort is lower when is also most valuable. The ability of the low skill agent to retain opportunities dramatically reduces the effectiveness of communication in spot markets. □

FIGURE I HERE

V. JOINT BILLING AND THE VALUE OF PARTNERSHIPS

As documented in our description of professional service firms in section II, joint billing allows agents to agree to refer opportunities to each other before they are actually realized, and to share their income regardless of who deals with them. We refer to such ex-ante contracts as partnerships. The economics literature has used this term to denote teams of agents sharing the revenue of a joint activity to which they all contribute. Our use, which is

more closely aligned to the legal term, refers to agents who share revenues and opportunities even when not all of them are eventually involved in production.

A. The partnership contract

In order to stress the similarities and differences with the previous case we first derive the corresponding incentive compatibility constraints and then we write the program that determines whether partnership arises or not in equilibrium.

First, constraint (6) requires that opportunity v_0 is not referred. The partnership prescribes effort level $\frac{y_l(v_0|v_0)}{\eta v_0}$:

$$s_l[y_l(v_0|v_0)] - \psi\left(\frac{y_l(v_0|v_0)}{\eta v_0}\right) \geq \overline{u}_l + s_l[y_h(v_0|v_1)], \tag{25}$$

where $y_h(v_0|v_1)$ is given by equation (9).[31]

It is worth emphasizing the difference between equation (25) and the corresponding incentive compatibility constraint in the spot market transaction (8). In the latter, the effort exerted by the low skill agent was always first best as he fully kept ownership rights over the opportunity drawn. In contrast, in the present situation the low skill agent pledges the draw, irrespective of it's quality, and the resulting output to the partnership. As we shall see, it may be in the best interest of the partnership to bias the effort applied by the low skill agent to the resolution of opportunity v_0 away from first best.

The contract must also ensure that the low skill agent prefers to refer a valuable opportunity v_1 rather than keeping it and pretend instead that v_0 was drawn.[32] Assume then that v_1 was drawn. The low skill agent may deviate by announcing v_0 and keeping the opportunity for himself. Clearly the resulting output must be consistent with this announcement and this can only be achieved if the effort supplied is $\frac{y_l(v_0|v_0)}{\eta v_1}$. To prevent such possible deviations the partnership must meet the following incentive compatibility constraint.

$$s_l[y_h(v_1|v_1)] + \overline{u}_l \geq s_l[y_l(v_0|v_0)] - \psi\left(\frac{y_l(v_0|v_0)}{\eta v_1}\right). \tag{26}$$

[31] To avoid introducing additional notation we use the same symbol $s_l[\cdot]$ and $s_l[\cdot]$ as in the previous section to denote the shares of output that the partnership awards the low and high skill agent respectively.

[32] Note that in the spot market case this was decided by the low skill agent, who was the residual claimant to the gains from a good allocation. This constraint is, as a consequence met by the signaling equilibrium contract when referrals exist as we show below.

Equation (26) should be compared with the rule that determines whether the referral takes place or not in the spot market case:

$$s_l[y_h(v_1|v_1)] + \overline{u}_l \geq \max_y \left\{ y - \psi\left(\frac{y}{\eta v_1}\right) \right\}. \tag{27}$$

The comparison between equations (26) and (27) holds the key to much of what follows. In the case of spot market transactions the alternative use of the valuable task is for the low skill agent to keep it, as seen in the right hand side of equation (27). In contrast, in the partnership the low skill agent does not have such a choice as the output has to be consistent with the announcement of opportunity v_0.

A judicious choice of $y_l(v_0|v_0)$ can relax the distortions on the valuable effort of the high skill agent that truthful communication entails. That is, it is by distorting $y_l(v_0|v_0)$ away from the first best, y_0^{fb}, that one obtains a value of $y_h(v_1|v_1)$ closer to y_1^{fb}. Clearly there are limits to the distortions on $y_l(v_0|v_0)$ as this output is also valuable and the pursuit of efficient communication does not come at the expense of fully destroying incentives for effort in the least valuable tasks.

Finally, and as in the case of the spot market contract, constraint (7) must hold, i.e. incentives have to be provided to the high skill agent in order to elicit effort from him on those tasks that are referred.

$$s_h[y_h(v_1|v_1)] - \psi\left(\frac{y_h(v_1|v_1)}{v_1}\right) \geq s_h[y_h(v_0|v_1)] - \psi\left(\frac{y_h(v_1|v_1)}{v_1}\left(\frac{v_0}{v_1}\right)\right). \tag{28}$$

We state next the contracting problem of the agents. The agents can always choose to forego the benefits of the partnership and take their opportunities, if necessary, to the spot market. Recall that the welfare level associated with the spot market, W^m, was given by expression (22).

Define first,

$$W^p(y_l, y_h) = \pi\left[y_l - \psi\left(\frac{y_l}{\eta v_0}\right) + \overline{u}_h\right] + (1-\pi)\left[y_h - \psi\left(\frac{y_h}{v_1}\right) + \overline{u}_l\right]. \tag{29}$$

Then the program $\mathcal{P}_p$ determines whether the contract arises or not in equilibrium.

Problem $\mathcal{P}_p$

$$\max_{partnership,market} \left\{ \max_{\{s_l[\cdot],s_h[\cdot],y_l(v_0|v_0),y_h(v_1|v_1)\}} W^p\left(y_l(v_0|v_0), y_h(v_1|v_1)\right) \quad , \quad \overline{W}^m \right\}$$

subject to

$$s_l[y_l(v_0|v_0)] - \psi\left(\frac{y_l(v_0|v_0)}{\eta v_0}\right) \geq \overline{u}_l + s_l\left[y_h(v_1|v_1)\frac{v_0}{v_1}\right] \tag{30}$$

$$\overline{u}_l + s_l[y_h(v_1|v_1)] \geq s_l[y_l(v_0|v_0)] - \psi\left(\frac{y_l(v_0|v_0)}{\eta v_1}\right) \tag{31}$$

$$s_h[y_h(v_1|v_1)] - \psi\left(\frac{y_h(v_1|v_1)}{v_1}\right) \geq s_h\left[y_h(v_1|v_1)\frac{v_0}{v_1}\right] - \psi\left(\frac{y_h(v_1|v_1)}{v_1}\left(\frac{v_0}{v_1}\right)\right) \tag{32}$$

$$s_l[y] + s_h[y] = y \quad \text{for} \quad y \in \mathcal{Y} \tag{33}$$

As in the spot market equilibrium, the contract will set $s_h[y]$ sufficiently low for those outputs that unambiguously identify the high skill agent as the one deviating from the prescribed action.

Consistently with the discussion in the introduction and section 3.2, the agents face no ex-post participation constraints in $\mathcal{P}_p$. There exists no limited liability or risk aversion on the agents side, nor any constraint on their ability to commit, so if the contract produces more utility than the spot market, they will enter into it. It is only the need to communicate truthfully the opportunities what leads to the distortions in the effort decisions of both agents.

Program $\mathcal{P}_p$ can be considerably simplified by proceeding as follows. First notice that if (30) is met with strict inequality then $s_l[y_h(v_0|v_1)]$ can be raised until (30) is met with equality, which does not modify the incentives of the low skill agent and can only improve those of the high skill agent (see equation (32)). A similar argument holds for equation (31). If it is met with strict inequality $s_l[y_h(v_0|v_0)]$ can be raised without distorting incentives. For this reason,

$$s_l[y_h(v_1|v_1)] = s_l[y_h(v_0|v_1)] + \Phi_l\left(y_l(v_0|v_0)\right), \tag{34}$$

and where

$$\Phi_l\left[y_l(v_0|v_0)\right] = \psi\left(\frac{y_l(v_0|v_0)}{\eta v_0}\right) - \psi\left(\frac{y_l(v_0|v_0)}{\eta v_1}\right). \tag{35}$$

$\Phi_l\left[y_l(v_0|v_0)\right]$ is the surplus foregone by the low skill agent when referring v_1 rather than keeping it and saving on effort to produce the prescribed output $y_l(v_0|v_0)$.

Plugging (34) in (32) yields

$$\Phi_h\left[y_h(v_1|v_1)\right] \geq \Phi_l\left[y_h(v_0|v_0)\right], \tag{36}$$

where

$$\Phi_h\left[y_h(v_1|v_1)\right] = y_h(v_1|v_1) - \psi\left(\frac{y_h(v_1|v_1)}{v_1}\right) - \left[y_h(v_1|v_1)\frac{v_0}{v_1} - \psi\left(\frac{y_h(v_1|v_1)}{v_1}\frac{v_0}{v_1}\right)\right]. \tag{37}$$

$\Phi_h\left[y_h(v_1|v_1)\right]$ is the surplus generated by the partnership when the high skill agent takes the prescribed action rather than the possible deviation. Incentive compatibility then requires that the allocation $[y_h(v_1|v_1), y_l(v_0|v_0)]$ is such that enough surplus is generated by the high skill agent to compensate the low skill agent for the surplus the latter foregoes when referring v_1.

The new program then reduces to,

Problem $\mathcal{P}'_p$

$$\max_{\{partnership,market\}}\left\{\max_{\{y_l(v_0|v_0),y_h(v_1|v_1)\}} W^p\left(y_l(v_0|v_0), y_h(v_1|v_1)\right) \quad , \quad \overline{W}^m\right\}$$

subject to

$$\Phi_h\left[y_h(v_1|v_1)\right] \geq \Phi_l\left[y_h(v_0|v_0)\right]$$

B. Characterization of the partnership contract

Some of the intuition of what follows can be gauged by looking at the constraint (36). Assume, for instance, that $\Phi_h[y_1^{fb}] < \Phi_l[y_0^{fb}]$. In this case the agents face the choice of distorting away from the first best either $y_h(v_1|v_1)$, $y_l(v_0|v_0)$, or both. The main difference with respect to the spot market case lies precisely in the possibility of worsening the incentives for effort in v_0 away from first best to relax the incentive compatibility constraint faced by the partnership. This possibility did not exist in the spot case because the market lacks the mechanism to enforce distortions on those activities that do not make it to the market in the first place. The introduction of joint billing achieves precisely this, namely, it allows agents to distort effort on activities that do not trigger team production to improve incentives for

communication. This in turn translates into better upstream incentives, where they are more valuable.

Clearly, output produced with opportunity v_0 is still valuable and the partnership weighs the gains from efficient communication and the improvements in upstream incentives against the incentives losses downstream. Notice that, in principle, the partnership can implement the first best communication rule by simply paying the low skill agent a fixed wage but in this case it can elicit no effort from her. The provision of incentives for effort interferes with optimal communication and for this reason, and as the next proposition shows, distortions may occur across all tasks, v_0 and v_1.

Proposition 5. The partnership allocation, $[y_0^p, y_1^p]$, is one of two types:

1. The first best can be implemented, that is, $[y_0^p, y_1^p] = \left[y_0^{fb}, y_1^{fb}\right]$.
2. If the first best cannot be implemented then $y_0^p < y_0^{fb} = y_0^m$ and $y_1^m < y_1^p < y_1^{fb}$.

To reiterate, effort exerted in v_0 is distorted away from it's first best level in order to improve incentives for communication. Clearly then the extent of the distortion depends on the frequency of communication that the first best prescribes, as summarized by the probability $1 - \pi$. As π increases communication becomes relatively less important than the preservation of the effort incentives on the tasks that are more frequent, v_0. The next proposition establishes the formal link between communication about economic opportunities and incentives for effort provision across those opportunities.

Proposition 6. y_0^p (y_1^p) is an increasing (decreasing) function of π.

C. Comparison with the spot market equilibrium

As shown in proposition 5, distortions away from first best may occur even in those opportunities that do not need to be communicated. Still the partnership does at least as well as the spot market.

Proposition 7. Any signaling equilibrium is a feasible allocation of the partnership.

The next corollary is then straightforward and is given without proof.

Corollary 8. (a) The partnership allocation Pareto dominates (weakly) the market allocation and

(b) The referral set supported by the signaling equilibrium is contained in the one supported by the partnership.

In the spot market signaling equilibrium the inefficiency sprang from the large reduction in high skill effort incentives that the optimal contracts entailed. As we proved here the ex-ante agreement taxes the low skill agent if she keeps the economic opportunity drawn in the first round. The low skill agent shares the proceeds of her effort with the high skill agent even when the latter is not involved in production. Clearly, this would never be an outcome of a spot market. It would involve a worker voluntarily, and in exchange of nothing, sending a share of the income he obtains to some skilled agent.

As a consequence of this tax, the low skill agent is willing to refer opportunities for a lower share and this improves the effort incentives of the high skill agent and enlarges the referral set. Recall though that in the spot market equilibrium the low skill agent is always performing first best effort in the opportunities he tackles. Instead, the partnership agreement may demand from *all* agents that they provide suboptimal effort. Still, the partnership improves on the market allocation because it trades off better effort incentives where these are most valuable for worse effort incentives where these are least valuable, namely on those tasks attempted by low skill agents.

This "re-allocation" of inefficiencies from one type of economic opportunities to the other is unique to the partnerships. By jointly billing the economic opportunities, they allow agents to commit to the contract and break the fundamental constraint of spot market transactions: that goods and services simultaneously flow in opposite direction to the spot price of those goods and services. It is precisely because the partnership distorts incentives in those opportunities that do not make use of all the resources of the organization that it can do better than the spot market.

We conclude this section by showing a simple example that compares the allocation associated with spot market transactions versus that of the partnership.

Example (cont.) We continue with the quadratic example introduced above. Trivial

computations show that

$$\Phi_h[y_1^{fb}] = \frac{(v_1 - v_0)^2}{2} \quad \text{and} \quad \Phi_l[y_0^{fb}] = \frac{(\eta v_0)^2}{2}\left[1 - \left(\frac{v_0}{v_1}\right)^2\right],$$

and hence restriction (36) translates into

$$\left(\frac{v_1}{v_0}\right)^2 \left[\frac{v_1 - v_0}{v_1 + v_0}\right] \geq \eta^2,$$

that is, if v_1 is sufficiently close to v_0 the ex-ante contract cannot implement the first best. As shown in figure II this occurs both in regions d and e. In region d the distortions are severe enough as to preclude communication completely, whereas in region e the partnership arises in equilibrium, but it cannot support the first best in equilibrium. When the distance between v_0 and v_1 is large enough the partnership arises to implement the first best.

The contrast with the referral set supported by the sharing contracts studied in the previous section is sharp (see figure I.) The partnership can implement the first best for v_1s that are high enough, where this is most valuable, whereas sharing contracts cannot even support their referrals. □

FIGURE II HERE

In summary both spot markets and partnerships can support the right flow of opportunities into the market but they do so at the expense of distortions in the effort provided by the agents involved in production. As we show in what follows, these distortions spring from the control exerted by the low skill agent on the information flow. When information flows downstream or horizontally there are no distortions associated with referrals.

VI. DOWNSTREAM REFERRALS, HORIZONTAL REFERRALS AND THE SCOPE OF THE PARTNERSHIP

The paper up to now has studied the case in which a less skilled agent observes an opportunity and must decide whether to refer it or not. In this section, we analyze two other cases, in both of which the informational asymmetry exists, but does not result in an incentive conflict. First, we study downstream referrals which are those in which a more skilled agent

has diagnosed and observed the value of the opportunity and must decide whether to refer it or not to a less skilled agent. Second, we study horizontal referrals, in which agents cannot threaten to withhold an opportunity that belongs to another agent and extract some value, as the skills of the agents are incompatible with each other's opportunities. In both cases, the same result holds: no rent-sharing contracts are required to ensure the 'right' referral flow.

A. Downstream referrals

Consider now the case in which an agent draws an opportunity for which he is overqualified. In this case, a referral market with no output sharing can achieve the first best allocation. Simply put, the higher skill agent prefers to refer the less valuable opportunities; thus a transfer of these opportunities downstream is incentive compatible from her perspective. Since this is the first best allocation, we can decentralize it even under informational asymmetries.

Proposition 9. If the opportunities flow from higher to lower skill agents, a fixed price contract supports the first best allocation.

B. Horizontal referrals

Consider the following minor variation of the vertical referrals model introduced in section III. As before one of the agents, say agent A, randomly draws opportunities that now come in two different "specialties" A and B, that is, $v \in \{v_A, v_B\}$. The production function is now,

$$y_i = \theta_i(v)ev \quad \text{for} \quad i \in \{A, B\},$$

where $\theta_i(v) = 1$ if $v = v_i$ and $\theta_i(v) = 0$ if $v \neq v_i$. Agent A can only tackle opportunities in field A whereas agent B can only do so in field B. Furthermore, and it order to guarantee trade assume that[33]

$$0 < \overline{u}_i < \max_e \{\theta_i(v_i)ev_i - \psi(e)\}. \tag{38}$$

Trivially then the first best prescribes matching agents of specialty i with v_i for $i \in \{A, B\}$. Even if the quality of the opportunity is not publicly observed there is no room for adverse

[33]The assumption that $\overline{u}_i > 0$ is made in order to avoid the unreasonable case where an agent may retain an opportunity of a different specialty to avoid the outside value.

selection problems to occur as it never pays to retain an opportunity of an specialty other than that of the agent in possession of the opportunity. As in the case of downstream referrals, fixed price transactions support the first best.[34]

Proposition 10. The first best horizontal referral flow can be supported by a fixed price contract.

C. Scope of the partnership: Discussion

Propositions 9 and 10 hold the last empirical implication of our model. The contractual structure is determined by the interaction between the relative skill of the agents and the control of the informational flow. If opportunities flow from less skilled to more skilled workers, giving a stake to less skilled workers in the correct communication is necessary in order to obtain the right allocation. As a result, partnerships contracts should be signed between agents who share some skills, and among whom this threat of appropriation binds.

On the other hand, when opportunities flow from more skilled to less skilled workers, or when opportunities flow between agents whose skills have no overlap in the sense that they could not choose to appropriate each other's opportunities, there is no need for an ex-ante referral arrangement. Empirically, this implies that agents who are always less qualified than any other agent should not receive revenue sharing. Moreover, and consistently with our discussion in section II.B, referrals between firms with different geographical or product scope should not involve sharing contracts.

In this sense, referrals between firms should take place. Professional services firms are, as in our model, based on ex-ante agreements to share the income derived from the opportunities obtained by each member. However, they do not require all opportunities to be referred in house. Law firms, for example, specialize in certain areas of the law and not in others. "Wall Street" law firms like Cravath or Cleary offer their clients legal advice on those areas that are the domain of the client's activities (securities law, mergers and acquisitions, commercial bank laws, etc,) but not in others (like maritime law.) If a client happens to have a problem in such a domain, they are referred to another firm that specializes in that field.[35]

[34] In particular, and as shown in the proof of proposition 10, zero price transactions support the first best allocation.

[35] A Cleary partner we were fortunate to interview claimed that the two most important reasons for the

This outcome is as predicted with our model. Agents should create ex-ante referral arrangements or partnerships when their skills overlap, so that a substantial threat of appropriability of the opportunity and the resulting misallocation exists. On the other hand, agents need not create a partnership when the opportunities belong to entirely different skill spaces (such as family law and corporate law) or entirely different geographical areas. Agents should team up in partnerships with those agents to whom they are likely to need to refer opportunities, but with whom the skills overlap sufficiently that the risk of misallocation is large.

Thus our theory generates a theory of the size and boundaries of the firm that supports transactions both inside the firm and in the market. The empirical implications about the types of practices that firms should include are very different from the implications of risk sharing theories, such as Gilson and Mnookin (1985). Rather than aiming to diversify the types of skills they include, partherships should aim to cover all areas which are sufficiently related that referrals are likely to be necessary, and where the threat that each agent can make of appropriating an opportunity is also high.

VII. CONCLUSIONS AND EXTENSIONS

The purpose of our analysis has been to study how economic organization assigns opportunities to those who are most qualified to deal with them in the presence of informational asymmetries about their value. When agents enjoy an informational advantage due to their knowledge of an opportunity they may be in fact not qualified to deal with, economic organization must provide them with incentives to both refer those opportunities they are not best qualified to deal with and to apply effort to those opportunities they keep. Moreover, incentives must be provided to those who receive the referral to work adequately on it.

We have obtained six main implications of our analysis. First, upstream referrals of opportunities are subject to adverse selection and, as a consequence, require that the referring agent take a stake in the output produced by the agent who receives the referral.

Second, truthful communication in the spot market comes at the expense of the comparative advantage of high skill agents. Lower (rather than higher) effort is applied the

referral to another law firm where the lack of expertise in a rare problem a client may have and conflicts of interest (when the same law firm represents two clients who have a pending suit or conflict.)

higher the value of the opportunity. For this reason communication is severely limited in spot market transactions.

Third, partnership contracts, in which agents commit to sharing the output of the opportunity even when it does not get transferred between agents, improve on the allocation of spot markets. The reason is that, by taxing a low skill agent who keeps an opportunity, they decrease her incentive to under-refer. The contracts trade-off the distortions they impose on both agents depending on the frequency with which opportunities must be referred. The higher the frequency of communication, the lower the output of the low skill agent and the higher the output of the high skill agent. Writing such complete contracts requires making the revenue from the opportunities observable and verifiable, regardless of who deals with it. This is the critical role played by joint billing of the clients by the partnership.

Fourth, the strength of the effort incentives is limited by the risk that the referring agent will prefer to appropriate the opportunity if given incentives which are too high powered. As a result, effort incentives are stronger the lower the ex-ante likelihood that a referral will be required to take place.

Fith, downstream and horizontal referrals do not require sharing of income between workers, as they are not subject to adverse selection, and, as a consequence, the first best obtains. Thus it is the relationship between the relative skill of the agents and the direction of the informational flow what determines the type of contracts observed.

Finally, these findings have implications for the scope of the partnership. In particular, we show that partnership contracts should appear between agents who are somewhat specialized, so that they have comparative advantages in different types of opportunities, but share the same skills to some extent. The reason is that it is among these agents that the threat of appropriation of the opportunities is more important. Partnerships arise to prevent the inefficiencies in the allocation that otherwise would occur.

There are a number of possible extensions of our analysis. First, we have assumed that the referring agent can either always appropriate the entire value of the opportunity she draws, so that her outside value is the value she could extract from the opportunity on her own; or, alternatively, the agent can, through joint billing, commit completely to share the income from the opportunity. Many relevant cases may however fall in between, where firms face severe difficulties in enforcing ex-ante agreements. For instance, partners in a law firm

may break up from the firm and take several of the assets, like clients and associates, with them. In this case a more general choice for the outside value of the low skill agent may be given by: $y = \delta\eta ev$.[36] We can interpret $1 - \delta$ as the productivity gain that results from the usage of the specific complementary assets of the long term organization. If the value of the opportunity is highly complementary to the firm's assets, the outside value is 0. The partnership knows that developing specific assets that tie employees to the firm, will improve the incentives for communication, as access to the specific assets of the firm reduces the value of defecting with an opportunity.[37] We leave for future research the in-depth study of these situations.

Future research also needs to study the implications of endogenizing the referral flow. An important implication of our set-up is the determination of the contract by the interaction between the relative skill of the agents and the direction of the flow of information. Intuitively, if the agents can design information flows, they will trade off informational asymmetries against information processing and specialization costs.

Also, in order to focus on the important informational asymmetries between agents, our model has ignored the role that the client may play in the referral process. Integrating the analysis of the relationship between the client and the first problem solving agent with the relationship between the two agents is an important issue for future work.

Finally, we have assumed that agents are risk neutral and have unlimited liability. Still, exogenous limits on the ability of the parties to execute transfers, either ex-ante or ex-post, also create distortions in effort provision. Future research may dispense with these assumptions. We believe, however, that in permitting agents to embark in all contracts that increase joint surplus our analysis has allowed us to focus on the trade-off between effort incentives and communication incentives. It is this trade-off rather than the 'tenuous trade-off' between risk and incentives (Prendergast (1999)) that plays in our view the key role in the contractual design of these professional service firms.

[36] Situations were the reservation utility is type dependent have been investigated by Lewis and Sappington (1989) and characterized more generally by Maggi and Rodriguez-Clare (1995) and Jullien (2000)

[37] Such an investment possibility by firms was first recognized by Rajan and Zingales (1998).

REFERENCES

Akerlof, George "The Market for Lemons: Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics,* LXXXIX (1970), 488-500.

Altonji, Joseph B. "Pay to Win: How to redeploy your firm's compensation system as a strategic asset." *The American Lawyer,* (February 2000).

Arrow, Kenneth J. Essays in the Theory of Risk Bearing. North-Holland New York, NY (1971).

Arrow, Kenneth J. The Limits of Organization. W.W. Norton & Company. New York, 1974.

Cotterman, James D. ed. Compensation Plans for Law Firms, 2nd ed., Altman, Weil, Pensa Inc. (1995), Newton Square, PA.

Demski, Joel S. and Sappington, David E. M. "Delegated Expertise." *Journal of Accounting Research* 25 (Spring 1987):68-89.

Farrell, Joseph and Scotchmer, Suzanne. "Partnerships." *Quarterly Journal of Economics*, Vol. 103, No. 2 (1988):279-298.

Garicano, Luis "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy* 108:4 (October 2000).

Gilson, R. J. and R.H. Mnookin, "Sharing Among the Human Capitalists: An Economic Inquiry into the Corporate Law Firm and how Partners Split Profits," *Stanford Law Review*, XXXVII (1985), 313-392.

Grossman, Sanford and Oliver Hart. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94, (1986), 691-719.

Hart, Oliver and John Moore. "Property Rights and the Nature of the Firm", *Journal of Political Economy*, 98 (1990), 1119-58.

Hayek, Friedrich A. "The Use of Knowledge in Society." *American Economic Review*, 35 (1945), 519-30.

Holmstrom, Bengt "Moral Hazard in Teams." *Bell Journal of Economics*, X (1982), 74-91.

Holmstrom, Bengt and Paul Milgrom "Multitask Principal-Agent Analysis: Incentive Contracts, Assets Ownership, and Job Design," *Journal of Law, Economics, and Organizations*, VII (1991), 24-52.

Hotelling, Harold "Stability in Competition," *Economic Journal*, 39 (1929), 41-57.

Jullien, Bruno, "Participation Constraints in Adverse Selection Models," *Journal of Economic Theory*, (2000).

Kehrer, Kenneth, "Trust vs. Brokerage: When both departments can handle the job, who gets the business?" *ABA Banking Journal*, October 1998, pp 95-98.

Kandel, Eugene and Lazear, Edward P. "Peer Pressure and Partnerships." *Journal of Political Economy,* vol. 100 N. 4 (1992): 801-817.

Laffont, Jean Jacques and Jean Tirole, A Theory of Incentives in Procurement and Regulation, MIT Press, 1993.

Legros, Patrick and Steven A. Matthews, "Efficient and Nearly-Efficient Partnerships," *The Review of Economic Studies*, Vol. 60, No. 3. (1993): 599-611.

Lewis, Tracy and David Sappington, "Countervailing Incentives in Agency Problems," *Journal of Economic Theory,* 49 (1989), 238-263.

Maggi, Giovanni and Andres Rodriguez-Clare, "On Countervailing Incentives," *Journal of Economic Theory,* 66 (1) (1995), 238-63.

Magnus, Stephen A. "Physicians' Financial Incentives in Five Dimensions: A Conceptual Framework for HMO managers," *Health Care Management Review*, 24(1) (1999), 57-72.

Marcus, Ruth, "Covington Challenge: To Stay on Top," *The National Law Journal*, Monday, May 4, 1981.

Mudrick, Howard, "Partner compensation." in The CPA Journal Online, at http://www. nysscpa.org /cpajournal/ old/08656260.htm, 1997.

Prendergast, Canice, "The Provision of Incentives in Firms," *Journal of Economic Literature*, XXXVII (1) (1999), 7-63.

Prendergast, Canice, "What Trade-off of Risk and Incentives?" *American Economic Review*, XC (2) (2000), 421-425.

Rajan, Raghuram and Luigi Zingales, "Power in the Theory of the Firm," *Quarterly Journal of Economics*, CXIII (1998), 387-432.

Rodwin, Marc A. Medicine, Money and Morals: Physicians' Conflicts of Interest. New York: Oxford University Press, 1993.

Rodwin, Marc A. "The Organized American Medical Profession's Response to Financial Conflicts of Interest: 1890-1992." *The Milbank Quarterly*, LXX(4) (1992), 703-741.

Rosen, Sherwin "Hedonic Prices and Implicit Markets," *Journal of Political Economy*, 82(1972), 34-55.

Spurr, Stephen J., "Referral Practices Among Lawyers: A Theoretical and Empirical Analysis." *Law and Social Inquiry.*(1988)

Spurr, Stephen J., "The Impact of Advertising and other Factors on Referral Practices, with Special Reference to Lawyers." *Rand Journal of Economics*, 21 no.2 (Summer 1990): 235-246.

Taylor, Curtis, "The Economics of Breakdowns, Checkups and Cures" *Journal of Political Economy* 103, no. 1 (February 1995):53-74.

Teece, David J. "Firm Organization, Industrial Structure and Technological Innovation", *Journal of Economic Behavior and Organization*, vol. 31 (1996):193:224.

Trotter, Michael, H. Profit and the Practice of Law, The University of Georgia Press, Athens and London, 1997.

Weingarten, Gene, "Breaking Up," *The National Law Journal,* Monday, June 1, 1981.

Wilber, James, "Partner Compensation Systems - How Firms Distribute Owner Profits," AltmanWeil, August (2000), downloadable at http://www.altmanweil.com/about/articles.

Wolinsky, Asher, "Competition in a Market for Informed Experts' Services." *Rand Journal of Economics* 24 (Autumn 1993):380-98.

APPENDIX

Proof of Proposition 1. Assume the price q_{as} supports the referral of v_1 and the non referral of v_0. Then it must be that

$$\max_e \{\eta e v_0 - \psi(e)\} > \overline{u}_l + q_{as} > \max_e \{\eta e v_1 - \psi(e)\}.$$

But $g_l(v) = \max_e \{\eta e v - \psi(e)\}$ is increasing in v, a contradiction. □

Signaling contracts

With a slight abuse of notation define the following function:

$$\Theta(y, \delta) = y\delta - \psi\left(\frac{y}{v_1}\delta\right) - \Delta\overline{u}, \tag{39}$$

where

$$\Delta\overline{u} = \overline{u}_h - \overline{u}_l. \tag{40}$$

Notice that the right hand side of expression (19) can then be written as $\Theta\left(y_h(v_1|v_1), \frac{v_0}{v_1}\right)$.

Notice that

$$\Theta\left(y_1^{fb}, \frac{v_0}{v_1}\right) = e_1^{fb} v_0 - \psi\left(e_1^{fb}\frac{v_0}{v_1}\right) - \Delta\overline{u}. \tag{41}$$

With some abuse of notation, let $U_l^r(y)$ the utility of the low skill agent if the referral of v_1 takes place as a function of y, that is

$$U_l^r(y) = y - \psi\left(\frac{y}{v_1}\right) - \Delta\overline{u}. \tag{42}$$

Clearly $U_l^r(y)$ is a concave function with a maximum at y_1^{fb}.

We establish in the next lemma some properties of these functions that prove useful below. The proof of these results is straightforward and is omitted.

Lemma A1. (a) $\Theta(y, \delta)$ is a strictly concave function of y with a maximum at $y = \frac{y_1^{fb}}{\delta}$.

(b) For $\frac{v_0}{v_1} \leq \delta \leq 1$

$$\Theta\left(y, \frac{v_0}{v_1}\right) < U_l^r(y) \quad \text{for} \quad y \leq y_1^{fb} \tag{43}$$

$$\Theta\left(y, \frac{v_0}{v_1}\right) > U_l^r(y) \quad \text{for} \quad y \geq y_1^{fb}\frac{v_1}{v_0} \tag{44}$$

Proof of Lemma A1. (a) It follows from $\psi''(\cdot) > 0$ and the definition of y_1^{fb}.

(b) To prove (43) notice that

$$\frac{d\Theta(y,\delta)}{d\delta} = y\left(1 - \frac{1}{v_1}\psi'\left(\frac{y}{v_1}\delta\right)\right) > y\left(1 - \frac{1}{v_1}\psi'\left(\frac{y_1^{fb}}{v_1}\right)\right) = 0,$$

where the strict inequality follows because $y \leq y_1^{fb}$ and $\delta \leq 1$. Because $U_l^r(y) = \Theta(y,1)$ the result follows. As for (44)

$$\frac{d\Theta(y,\delta)}{d\delta} = y\left(1 - \frac{1}{v_1}\psi'\left(\frac{y}{v_1}\delta\right)\right) < y\left(1 - \frac{1}{v_1}\psi'\left(\frac{y_1^{fb}}{v_1}\right)\right) = 0,$$

as $y \geq y_1^{fb}\frac{v_1}{v_0}$ and $\delta \geq \frac{v_0}{v_1}$. Because $U_l^r(y) = \Theta(y,1)$ the result follows. □

Proof of Proposition 2. Clearly the low skill agent always performs first best effort in v_0, and $y_0^m = y_0^{fb}$.

Assume first that task v_1 is indeed referred. The low skill agent can extract all the surplus from the transaction and hence, if the first best can be implemented, he will indeed offer a contract to support it and that leaves him with all the rents.

Assume next that the first best cannot be implemented. Notice that $\Theta(0, \frac{v_0}{v_1}) = U_l^r(0) = -\Delta\overline{u}$. Furthermore $\Theta\left(y_1^{fb}\left(\frac{v_1}{v_0}\right), \frac{v_0}{v_1}\right) = U_l^r(y_1^{fb})$, that is, both function attain the same value at their respective maxima. By assumption (2)

$$\Theta\left(y_1^{fb}\left(\frac{v_1}{v_0}\right), \frac{v_0}{v_1}\right) > y_0^{fb} - \psi\left(\frac{y_0^{fb}}{\eta v_0}\right).$$

As shown in Lemma A1-(a), $\Theta\left(y, \frac{v_0}{v_1}\right)$ is strictly concave and hence it intersects $y_0^{fb} - \psi\left(\frac{y_0^{fb}}{\eta v_0}\right)$ twice and only twice, at $y = y_1^{m_a}$ and $y = y_1^{m_b}$, where $y_1^{m_a} < y_1^{m_b}$.

Clearly $y_1^{m_a} < y_1^{fb} < y_1^{m_b}$, otherwise the first best could be implemented. Given the production technology $e_h^{s_a} < e_h^{fb} < e_h^{s_b}$ and the signaling equilibrium s_a is associated with the underprovision of effort and s_b is associated with the overprovision of effort.

To show that the signaling equilibrium s_a Pareto dominates s_b, notice first that the high skill agent enjoys utility $\overline{u}_h$ in either equilibrium. Hence the only consideration is whether the low skill agent is better treated in one versus the other.

Given that $y_1^{m_a} < y_1^{fb} < y_1^{m_b}$ it follows from (43) and (44) that

$$U_l^r\left(y_1^{m_a}\right) > \Theta\left(y_1^{m_a}, \frac{v_0}{v_1}\right) = \Theta\left(y_1^{m_b}, \frac{v_0}{v_1}\right) > U_l^r\left(y_1^{m_b}\right),$$

and s_a Pareto dominates s_b.

Finally if the distortions are severe enough as to prevent the referral altogether then the out produced by the low skill agent with opportunity v_1 is given by $\eta\widetilde{e}v_1$ where $\widetilde{e} = argmax\{\eta e v_1 - \psi(e)\}$, but clearly $\widetilde{e}$ is an increasing function of η and hence $\eta\widetilde{e}v_1 < y_1^{fb}$. □

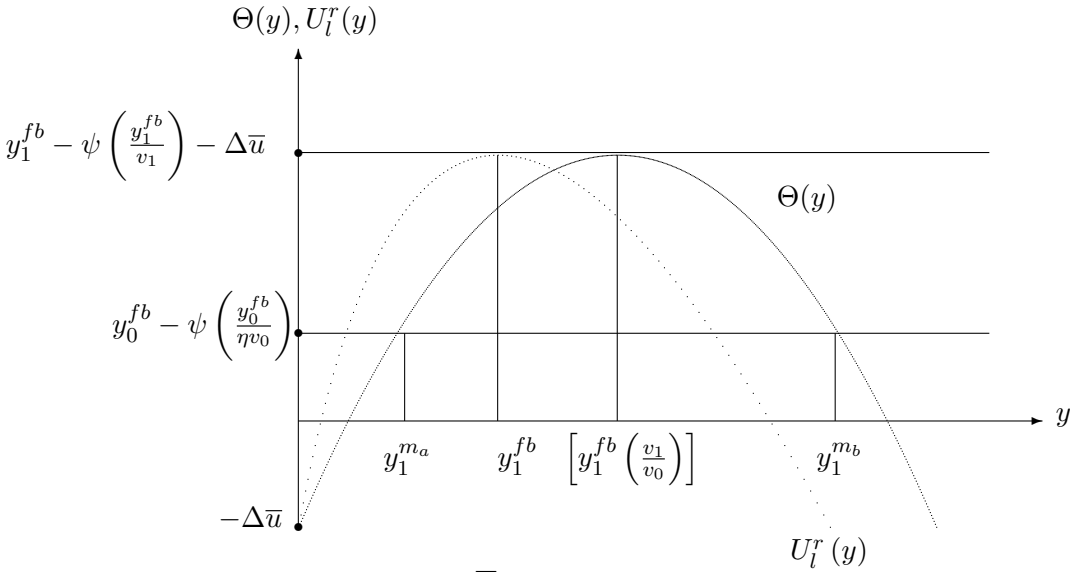

FIGURE III
Proposition 2

Proof of Proposition 3. If the first best cannot be implemented the effort of the high skill agent, e_1^s, is determined by:

$$y_0^{fb} - \psi\left(\frac{y_0^{fb}}{\eta v_0}\right) = e_1^s v_0 - \psi\left(e_1^s \frac{v_0}{v_1}\right) - \Delta\overline{u},$$

and applying the implicit function theorem,

$$\frac{\partial e_1^s}{\partial v_1} = -\frac{e_1^s}{v_1^2}\left(\frac{\psi\left(e_1^s \frac{v_0}{v_1}\right)}{1 - \frac{1}{v_1}\psi\left(e_1^s \frac{v_0}{v_1}\right)}\right) < 0,$$

which concludes the proof. □

Proof of Corollary 4. We proceed by contradiction. Assume that v_1 is such that (2) is not met but that

$$u_l^r(v_1) \geq \max_y \left\{y - \psi\left(\frac{y}{\eta v_1}\right)\right\},$$

that is, such that the signaling equilibrium supports the referral of v_1 whereas the first best policy does not. Because condition (2) is violated the first best level of welfare is

$$\overline{W}^{fb} = \overline{u}_h + \pi \max_y \left\{y - \psi\left(\frac{y}{\eta v_0}\right)\right\} + (1-\pi)\max_y \left\{y - \psi\left(\frac{y}{\eta v_1}\right)\right\},$$

whereas that of the signaling equilibrium is

$$W^m = \overline{u}_h + \pi \max_y \left\{ y - \psi\left(\frac{y}{\eta v_0}\right)\right\} + (1-\pi)u_l^r(v_1).$$

But $\overline{W}^{fb} > W^m$ which implies that $u_l^r(v_1) < \max_y \left\{ y - \psi\left(\frac{y}{\eta v_1}\right)\right\}$, a contradiction. □

Partnership contracts

Let $W[y_1, y_0]$ be the welfare function defined in the space of output allocations resulting from task v_1 and v_0 respectively.

Lemma A2. $W[\cdot,\cdot]$ is quasiconcave.

Proof: Recall that the welfare function is:

$$W[y_1, y_0] = \pi\left[y_0 - \psi\left(\frac{y_0}{\eta v_0}\right) + \overline{u}_h\right] + (1-\pi)\left[y_1 - \psi\left(\frac{y_1}{v_1}\right) + \overline{u}_l\right]$$

Let $[y_1^a, y_0^a]$ and $[y_1^b, y_0^b]$ be two allocations such that $W[y_1^a, y_0^a] = W[y_1^b, y_0^b] = \overline{W}$ Form a convex combination of both allocations, $[\hat{y}_1, \hat{y}_0] = \alpha[y_1^a, y_0^a] + (1-\alpha)[y_1^b, y_0^b]$. Recall that $\psi(\cdot)$ is a convex function. Then:

$$\psi\left(\frac{\alpha y_0^a + (1-\alpha)y_0^b}{\eta v_0}\right) < \alpha\psi\left(\frac{y_0^a}{\eta v_0}\right) + (1-\alpha)\psi\left(\frac{y_0^b}{\eta v_0}\right),$$

and similarly with $\hat{y}_1$. Hence:

$$\begin{aligned} W[\hat{y}_1, \hat{y}_0] &> \pi\left[\alpha\left(y_0^a - \psi\left(\frac{y_0^a}{\eta v_0}\right)\right) + (1-\alpha)\left(y_0^b - \psi\left(\frac{y_0^b}{\eta v_0}\right)\right)\right] \\ &+ (1-\pi)\left[\alpha\left(y_1^a - \psi\left(\frac{y_1^a}{v_1}\right)\right) + (1-\alpha)\left(y_1^b - \psi\left(\frac{y_1^b}{v_1}\right)\right)\right] \\ &> \overline{W} \end{aligned}$$

This concludes the proof. □

A straightforward application of the implicit function theorem results in the following slope for the indifference curve:

$$\frac{dy_0}{dy_1} = -\left(\frac{1-\pi}{\pi}\right)\left[\frac{1 - \frac{1}{v_1}\psi'\left(\frac{y_1}{v_1}\right)}{1 - \frac{1}{\eta v_0}\psi'\left(\frac{y_0}{\eta v_0}\right)}\right] < 0 \quad \text{for all} \quad y_0 < y_0^{fb} \quad \text{and} \quad y_1 < y_1^{fb} \tag{45}$$

Notice that as:

$$y_1 \to y_1^{fb} \quad \text{then} \quad \frac{dy_0}{dy_1} \to 0 \tag{46}$$

$$y_0 \to y_0^{fb} \quad \text{then} \quad \frac{dy_0}{dy_1} \to -\infty \tag{47}$$

Clearly, the indifference map has a satiation point in $\left[y_1^{fb}, y_0^{fb}\right]$ and bend "backwards" whenever $y_1 > y_1^{fb}$ or $y_0 > y_0^{fb}$.

Define the

$$\mathcal{S}^c = \{[y_1, y_0] \quad \text{such that} \quad \Phi_h(y_1) \geq \Phi_l(y_0)\},$$

that is, the set of incentive compatible allocations, and define $\mathcal{F}^c$ as the frontier of $\mathcal{S}^c$, which defines y_0 as a function of y_1.

Lemma A3 The frontier of incentive compatible contracts is a strictly concave function with a maximum at $y_1^* < y_1^{fb}$.

Proof: A straightforward application of the implicit function theorem shows that,

$$\left[\frac{dy_0}{dy_1}\right]_{\mathcal{F}^c} = \frac{1 - \frac{1}{v_1}\psi'\left(\frac{y_1}{v_1}\right) - \frac{v_0}{v_1}\left[1 - \frac{1}{v_1}\psi'\left(\frac{y_1}{v_1}\frac{v_0}{v_1}\right)\right]}{\Phi'(y_0)}. \tag{48}$$

Rearranging the above expression and applying the implicit function theorem again:

$$\left[\frac{dy_0}{dy_1}\right]_{\mathcal{F}^c} = -\frac{\left[\frac{d^2y_0}{dy_1^2}\right]_{\mathcal{F}^c}\Phi'(y_0) + \frac{1}{v_1^2}\left[\psi''\left(\frac{y_1}{v_1}\right) - \left(\frac{v_0}{v_1}\right)^2\psi''\left(\frac{y_1}{v_1}\frac{v_0}{v_1}\right)\right]}{\left[\frac{dy_0}{dy_1}\right]_{\mathcal{F}^c}\Phi''(y_0)}$$

or more clearly:

$$1 = -\frac{\left[\frac{d^2y_0}{dy_1^2}\right]_{\mathcal{F}^c}\Phi'(y_0) + \frac{1}{v_1^2}\left[\psi''\left(\frac{y_1}{v_1}\right) - \left(\frac{v_0}{v_1}\right)^2\psi''\left(\frac{y_1}{v_1}\frac{v_0}{v_1}\right)\right]}{\left(\left[\frac{dy_0}{dy_1}\right]_{\mathcal{F}^c}\right)^2\Phi''(y_0)}$$

which necessarily implies:

$$\left[\frac{d^2y_0}{dy_1^2}\right]_{\mathcal{F}^c} < 0,$$

as recall that $\psi'''(\cdot) \geq 0$,[38] $\Phi'(\cdot) > 0$ and $\Phi''(\cdot) > 0$.

Finally because $\Phi_h(y)$ is a strictly concave function of output it has a unique maximum at y_1^*. It follows from the first order conditions that $y_1^* < y_1^{fb}$. □

Proof of Proposition 5. The only interesting situation is when $\left[y_1^{fb}, y_0^{fb}\right]$ does not belong to $\mathcal{S}^c$. By Lemma A3 the frontier is a strictly concave function (the set of incentive compatible allocations $\mathcal{S}^c$ is convex.) Furthermore $\mathcal{S}^c$ is a compact set. Maximization of the strictly quasiconcave welfare

[38] This is the only step where this assumption is used.

function on the feasible set yields a unique maximum. Clearly, the planner places the allocation in the downward sloping side of the frontier (see figure III). The slope of the frontier of incentive compatible allocations, as given by (48), evaluated at either y_1^{fb} or y_0^{fb} is $-\infty < \left[\frac{dy_0}{dy_1}\right]_{\mathcal{F}^c} < 0$. But then by (46) and (47), it pays to move the allocation towards a strict interior. □

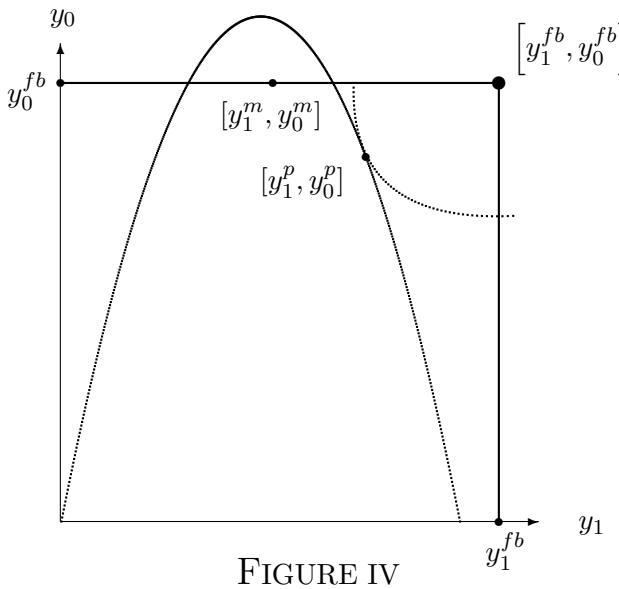

FIGURE IV
Proposition 5

Proof of Proposition 6. Clearly if the organization can implement the first best or if it cannot support any communication whatsoever then both y_1^p and y_0^p are independent of π.

If, on the other hand, the organization can only support communication with distortions then (36) is binding. The derivative of y_0^p with respect to y_1^p, as given by equation (48), is negative in the efficient side of the frontier of incentive compatible contracts (see figure IV). Taking the derivative of W^p, as given by equation (29), with respect to y_1^p, to obtain the first order condition and applying the implicit function theorem to find $\frac{dy_1^p}{d\pi}$ yields the result.□

Proof of Proposition 7. Trivially if the signaling equilibrium cannot support any referrals, then the utilitarian planner can always support a non referral rule by setting the transfers to the low skill agent low enough, possibly negative, conditional on any referral taking place.

Let $[y_0^{fb}, y_1^m]$ the equilibrium allocation when they are referrals in the signaling equilibrium, that is, whenever,

$$y_1^m - \psi\left(\frac{y_1^m}{v_1}\right) - \Delta\overline{u} \geq \max_y \left\{ y - \psi\left(\frac{y}{\eta v_1}\right) \right\}, \tag{49}$$

where the left hand side is the referral utility of the low skill agent as given by (42). We have to find transfers $s_l^p[y_0^{fb}]$, $s_l^p[y_1^m]$ and $s_l^p[y_1^{fb}\frac{v_0}{v_1}]$ such that the signaling allocation can be supported as a

feasible allocation of the partnership. Let

$$(50) \qquad s_l^p[y_0^{fb}] = y_0^{fb}, \quad s_l^p[y_1^m] = y_1^m - \psi\left(\frac{y_1^m}{v_1}\right) - \overline{u}_h, \quad \text{and} \quad s_l^p\left[y_1^m \frac{v_0}{v_1}\right] = s_l^s\left[y_1^m \frac{v_0}{v_1}\right],$$

where $s_l^s[y_1^m \frac{v_0}{v_1}]$ is the share of output that accrues to the low skill agent in the event of of the realization of output $y_1^m \frac{v_0}{v_1}$.

First (30) is trivially met as once we substitute $s_l^p[y_0^{fb}]$ by y_0^{fb} and $s_l^p[y_1^m \frac{v_0}{v_1}]$ by $s_l^s[y_1^m \frac{v_0}{v_1}]$, one immediately obtains (12). As for (31), using (50) and (49), yields,

$$\begin{aligned}
\overline{u}_l + s_l^p[y_1^m] &= y_1^m - \psi\left(\frac{y^m}{v_1}\right) - \Delta\overline{u} \\
&\geq \max_y \left\{ y - \psi\left(\frac{y}{\eta v_1}\right) \right\} \\
&> y_0^{fb} - \psi\left(\frac{y_0^{fb}}{\eta v_1}\right) \\
&= s_l^p[y_0^{fb}] - \psi\left(\frac{y_0^{fb}}{\eta v_1}\right)
\end{aligned}$$

As for (32), start by recalling that equation (14) is met with equality by market clearing (see equation (16)) and that (15) holds. Then:

$$\begin{aligned}
s_h^p[y_1^m] - \psi\left(\frac{y_1^m}{v_1}\right) &= y_1^m - s_l^p[y_1^m] - \psi\left(\frac{y_1^m}{v_1}\right) \\
&= \overline{u}_h \\
&\geq s_h^s\left[y_1^m \frac{v_0}{v_1}\right] - \psi\left(\frac{y_1^m}{v_1}\frac{v_0}{v_1}\right) \\
&= s_h^p\left[y_1^m \frac{v_0}{v_1}\right] - \psi\left(\frac{y_1^m}{v_1}\frac{v_0}{v_1}\right),
\end{aligned}$$

where the inequality follows from (13). □

Proof of Proposition 9. The contract consists of the high skill agent referring a low value opportunity v_0 for a fixed price q. Incentive compatibility requires that:

$$(51) \qquad \max_e \{ev_1 - \psi(e)\} \geq \overline{u}_h + q \quad \geq \quad \max_e \{ev_0 - \psi(e)\}$$

$$(52) \qquad \max_e \{\eta e v_0 - \psi(e)\} - q \quad \geq \quad \overline{u}_l.$$

Since there are more low skill agents competing for low value opportunities, they are kept at their reservation values, so that from (52), $q = \max_e \{\eta e v_0 - \psi(e)\} - \overline{u}_l$, and condition (52) is trivially met.

As for (51) first notice that,

$$\overline{u}_h + q = \overline{u}_h + \max_e \{\eta e v_0 - \psi(e)\} - \overline{u}_l > \max_e \{e v_0 - \psi(e)\},$$

where the last inequality follows from (1). Next notice that,

$$\begin{aligned} \overline{u}_h + q &= \overline{u}_h + \max_e \{\eta e v_0 - \psi(e)\} - \overline{u}_l \\ &< \overline{u}_h + \max_e \{\eta e v_1 - \psi(e)\} - \overline{u}_l \\ &< \max_e \{e v_1 - \psi(e)\}, \end{aligned}$$

where, once again, the last inequality follows from (2). Hence condition (51) is met. □

Proof of Proposition 10. Let the fixed price of transacted claims be q. We show that any price q such that

$$0 \leq q \leq \min \left[\max_e \{\theta_A(v_A) e v_A - \psi(e)\} - \overline{u}_A, \max_e \{\theta_B(v_B) e v_B - \psi(e)\} - \overline{u}_B \right] \tag{53}$$

supports the first best horizontal referral flow. Without loss of generality assume that an agent of specialty A draws a problem of specialty B. Clearly he prefers to refer it as he obtains $\overline{u}_A + q$. As for the receiving agent, that of specialty B, he accepts the referral as (53) guarantees that the price q of the claim is low enough to compensate him for the opportunity cost $\overline{u}_B$. Assume next that agent A draws a claim on his own specialty. If he falsely refers it as claim B he stands to obtain $\overline{u}_A + q$ whereas if he keeps it he earns $\max_e \{\theta_A(v_A) e v_A - \psi(e)\}$. But by (53), $\max_e \{\theta_A(v_A) e v_A - \psi(e)\} \geq \overline{u}_A + q$. Price q then supports the referral of claims B and the non-referral of claim A, as the first best prescribed. The proof is symmetric for the agent of specialty B. □

Signaling through linear sharing rules

The reader may be skeptical about the informational requirements of a solution like the one suggested before, since the contracts prescribe different sharing rules depending on whether the output is the equilibrium output or out of equilibrium. Cognitive limitations by part of agents may also bias the contracts towards simpler forms.

Here we briefly show that a separating signaling equilibrium can also be obtained in the spot market when we restrict the contracting space to two part tariffs. That is, the contract awards the low skill agent $s_l[y] = p + s \cdot y(v_1|v_1)$, where p is the fixed price and s is the portion of output obtained by the low skill agent when she refers the task, which is independent of the actual output produced.

The incentive compatibility constraint associated with the low skill agent whose draw is v_0 is now:

$$y_0^{fb} - \psi\left(\frac{y_0^{fb}}{\eta v_0}\right) \geq \overline{u}_l + p + s y(v_1|v_1)\left(\frac{v_0}{v_1}\right). \tag{54}$$

As for the high skill agent he chooses his effort level to maximize his utility. Recall that now the share of output that corresponds to the high skill agent is $s_h[y(v_1|v_1)] = (1-s)\cdot y(v_1|v_1) - p$. Then the output produced by the high skill agent will be the solution to:

$$(1-s)v_1 = \psi'\left(\frac{y(v_1|v_1)}{v_1}\right), \tag{55}$$

and the participation constraint of the high skill agent is now

$$(1-s)y(v_1|v_1) - p - \psi\left(\frac{y(v_1|v_1)}{v_1}\right) \geq \overline{u}_h \tag{56}$$

which, given market clearing, leads to $p = (1-s)y(v_1|v_1) - \psi\left(\frac{y(v_1|v_1)}{v_1}\right) - \overline{u}_h$.

The problem of the low skill agent is now to maximize (11) subject to (54), (55), and the participation constraint (56).

Note that, unlike before, the linear sharing contract always distorts the output decision of the high skill agent with respect to the first best.

The following lemma, whose proof is along similar lines as proposition 2, shows that there exists a linear sharing contract that can support separation.

Lemma A4. There exists a two part tariff (p^*, s^*) with $s^* \in (0,1)$ which supports a separating signaling equilibrium. This solution imposes larger distortions than the optimal spot market signaling contract and has less referrals.

Thus simpler, two part tariff, contracts can achieve the exact same aim of ensuring that the opportunities referred are the more valuable ones, but the linear restriction imposes larger distortions on output and, as a consequence, it supports referrals on an ever narrower set of opportunities.

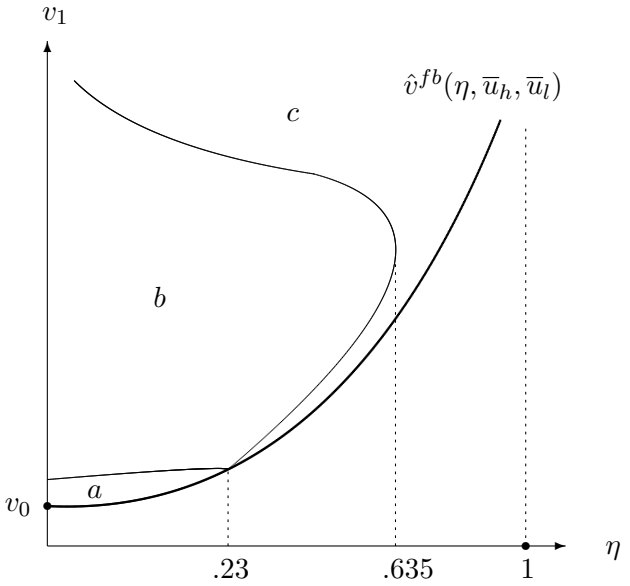

FIGURE I

Region a denotes the set of economies for which the spot market implements the first best. Region b denotes the set of economies for which the spot contract supports communications with distortions. Region c is the set of economies for which the spot market cannot support any communication. Recall that the first best prescribes the referral of any $v_1 > \hat{v}^{fb}(\eta, \overline{u}_h, \overline{u}_l)$ *and the non referral of any* $v_0 \leq \hat{v}^{fb}(\eta, \overline{u}_h, \overline{u}_l)$.

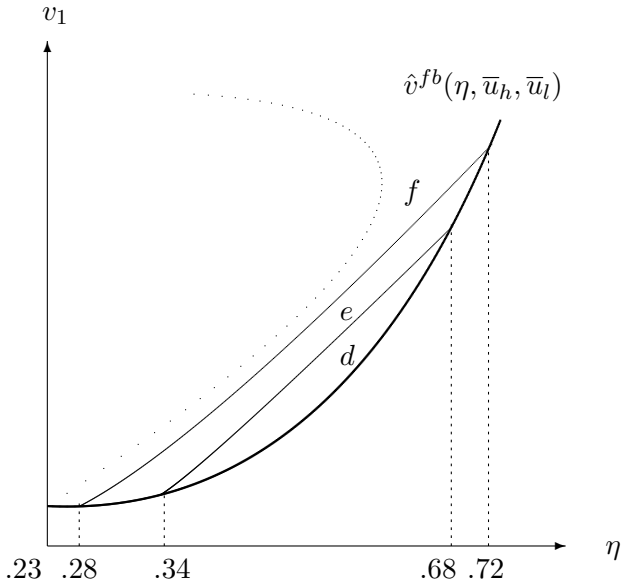

FIGURE II

Region d denotes the set of economies for which organizations do not arise. Region e denotes the set of economies where the organization supports communications but with distortions. Region f denotes the set of economies for which the organization implements the first best. The dotted line denotes the frontier for the spot market case.