Optimal contracting and the organization of knowledge*

William Fuchs Luis Garicano Luis Rayo
Berkeley-Haas LSE LSE
Preliminary, revision in course.
November 27, 2013

Abstract

We study which contractual arrangements may support the optimal organization of knowledge. Knowledge markets are plagued by informational problems, since both the difficulty of the questions posed and the skill of those advising on their answer are often hard to assess. We show that competitive spot contracting between producers and consultants suffers from excess entry as the lemons (in this market, intermediately skilled agents) cannot be excluded. An ex ante, firm-like contract, involving a full residual claim for the consultants and an increasing bonus payment for the producers is the unique contract that delivers first best whenever the spot contract suffers from excess entry. This simple characterization of the optimal ex ante contract suggests a rationale for the organization of firms and the structure of compensation in the knowledge intensive sector.

I Introduction

A key role of markets and other organizations is ensuring that society uses optimally the available knowledge (Hayek, 1945). Doing so requires, in particular, that the superior knowledge of experts be conserved for the ‘right’ questions: ensuring the correct matching of questions and expertise. A solution to this problem is a ‘knowledge-based hierarchy’ which allows experts to leverage their expertise by having less expensive workers deal with routine tasks.\(^1\) Firms and other organizations often structure these hierarchies by placing agents in different positions according to their expertise.

The allocation of talent is complicated by the presence of information asymmetries. First, the actual difficulty of the question posed is often hard to assess. Second, the skill of a consultant may also be unobservable. As a result, spot contracting is unlikely to be efficient. Indeed, inefficiencies

\(^{*}\)Garicano would like to thank the Toulouse Network in Information Technology for financial support. We thank participants in seminars at the AEA, Chicago, Boston U, LSE, Carlos III, CEMFI, and UC Berkeley for their comments. A previous version of this paper was circulated under the title “Trading Know-How.”

\(^{1}\)Such knowledge-based hierarchies, in which homogeneous workers must acquire knowledge and choose occupations under full information, where introduced in Garicano (2000). Garicano and Rossi-Hansberg (2006) analyze the problem of knowledge acquisition and optimal organizational structure with heterogeneous agents, also under full information.
can arise from too little or too much trade, and from an inefficient matching between problems and experts. In this paper, we study whether there exist contracts that can support an efficient allocation of knowledge despite these informational asymmetries, and if so what form such contracts must take.

We carry out our analysis in a general equilibrium setting, in which agents, each endowed with a (privately observed) skill level and a unit of time, endogenous choose their occupation (consultants or producers) and whether they wish to participate in the market for knowledge. Problems are encountered in the course of production, and vary in their difficulty. If a producer has a problem that is too hard relative to his skill level, he may seek help from a consultant.

We show that ex post markets (where agents who know they need help ask for it) cannot, in general, achieve efficiency, as too many ‘hard’ questions and too many ‘unqualified’ consultants appear in equilibrium. Instead, ex ante contracts, where agents pair with each other under asymmetric information about their types, but before they know which problems will actually require advice, always achieve the first best allocation. In fact, we show that the unique optimal contracts supporting efficiency have a quasi-firm feel to them, with non experts receiving increasing bonuses for solving their problem and experts receiving the full value of the problems on which they are asked for help.

We start by considering, as a benchmark, the optimal allocation when information asymmetries are absent. The more talented agents specialize in production and the more talented ones becoming specialized consultants. Furthermore, the first best involves positive sorting: the best consultants tackle the problems that are expected to be more difficult, which are those that the most skilled production agents could not solve by themselves. Intermediate agents do not enter the market. They are not talented enough to offer to solve others’ problems, but if they elect to ask for help, they pose questions that are, in expectation, too difficult to solve. The existence of these agents ‘in the middle’ is, as we shall see, the key obstacle for the development of a consulting market and is one of the features that distinguishes this type of market from a generic asymmetric information setting: the ‘lemons’ are the agents in the middle, and they are the ones which must be excluded from the market. Thus team production benefits most the most talented agents (who can specialize in harder problems for which they have a comparative advantage) and the least talented agents (who can specialize in production and get help from others). Those in the middle benefit least.

The main objective of our analysis is to study contracting under double-sided asymmetric information, that is when neither producers nor consultants know the talent of those they are matched with. We study two main possible timings. In ex post contracting (i.e. akin to the consulting markets we observe), producers seek consulting help whenever they face a problem they cannot solve. In the ex ante contracts, by contrast, consultants and producers pair up before producers know if they can solve or not the problems they face (in a similar way as, for example, inside a firm).

We show that a necessary condition for efficiency in either ex ante or ex post contracting is that
consultants must earn the full residual for the problems they consult on. As we show, the highest types of consultants are matched with the marginal intermediate type and thus must earn the full rents of the match, while the lowest types must be indifferent between staying out of the market and being consultants and thus must earn no rents. The only way this is possible is if all the consultants are earning the full marginal social value of their talent.

Given this result, the spot market features a special kind of inefficiency: too much trade, as no one can be excluded from the market. In the spot contracts, the best producers (who should be excluded) have incentives to pretend that their problems are simpler than they in fact are, in an effort to be able to get help on their problems. At the same time, the best producers have an incentive to pretend that they are more skilled than they are, in order to earn a fee as consultants. In other words, as consultants, they want to play smart, whereas as originators they want to play dumb. Excluding agents from the market requires both that asking for advice is costly (to avoid the independents playing dumb) but at the same time that agents pay for problems (to avoid independents playing smart). Solving this dilemma requires taxing the producers upfront, when they keep their his opportunity.

Our final results, and the key ones in the paper, are more promising for efficiency. We show that the first best can be attained, in general, when ex ante contracting is possible. Ex-ante contracting provides an additional instrument to support independent producers: when a consultant taxes a producer for the problems that this producer manages to solve, producers with high types (those who would optimally remain outside of the market) have a new incentive to stay away.

In fact, we show that in the optimal contract, the consultant is the residual claimant, while the producer receives conditional payments for solving problems. This simple, firm-like contract can be shown to always attain the first best, and to be the unique one that does so when independents must exist in equilibrium. The contracts have indeed a simple structure: the conditional payment for the expert when he solves the problem is 100% of the output, with a zero share for the originating agent, while the producer receives an increasing share (or is taxed at a decreasing rate), finishing in receiving the full output when she solves the problem.

No previous study has, to our knowledge, examined the double-sided adverse selection issue raised by matching consultants to problems under asymmetric information. The previous literature on consultant services emphasized moral hazard issues involved in the provision of consultant services, i.e., consultants have little incentive to supply the appropriate level of effort. Demski and Sappington (1987) examine the trade-off between productive effort and information gathering incentives faced by the consultant. In Wolinsky (1993), the issue is the incentives that must be provided for consultants to recommend the correct treatment, i.e., minor treatment for minor problems and major treatment for major problems. Wolinsky shows that specialization is optimal in this case. Similarly, Pesendorfer and Wolinsky (2003) study the provision of adequate diagnosis effort by consultants. Taylor (1995) studies how insurance can solve informational asymmetries in a context where only
the consultant can determine the necessary treatment. Garicano and Santos (2004) study the incentives for agents to pass on opportunities they could keep in a context with moral hazard and one-sided adverse selection. We depart from this literature in that we abstract from moral hazard and focus both on the on the double-sided nature of the informational asymmetries: the agent does not know the quality of the consultant, and the consultant does not know, a priori, the difficulty of the problem posed.\(^2\)

Our paper also fits within a literature that studies trade in markets with bilateral asymmetric information. Most of the literature stems from the original analysis by Myerson and Satherwaite (1983) of trade mechanisms under asymmetric information which addresses multiple buyers and sellers of a commodity with unknown valuation (e.g., Lu and Robert, 2001). In this literature, buyers and sellers do not care about each other’s quality per se, as they only care about the value of the object at stake. Thus, matching is irrelevant. The only paper we are aware of that studies equilibria in matching markets with two-sided incomplete information is Gale (2001). There are several important differences between our models. First, Gale takes as exogenous the side of the market in which agents are, while in our model, agents select ex-into buyers or sellers of advice. This endogenous choice makes the analysis somewhat more difficult, but the indifference conditions of the cutoff types help us pin down the equilibrium. Additionally, in Gale’s paper, all agents have equal outside option, while in our setting higher quality agents have a higher outside option, which increases the adverse selection problems because those agents exit the market first if it is unattractive. Finally, because our problem has more structure, we are able to go substantially further in characterizing the market equilibrium.

The paper is also related to the management-worker sorting literature and in particular Garicano and Rossi-Hansberg (2004 and 2006). As in those papers, the economy-wide problem studied here is one of matching talent with problems. However, here we study the problem under informational asymmetries absent from those papers and thus focus on optimal contracting. The worker/manager sorting models have been studied and generalized by Eeckhout and Kircher (2013), who study the general conditions under which interactions in production between worker and managerial skills generate skill-scale effects and positive sorting effects.

II Model set up

Individuals and skills. The economy is formed by a continuum of agents endowed with one unit of time and an ability level \(z \in [0, 1]\). Abilities are private information. We normalize skill so that an agent with skill \(z\) can solve a fraction \(z\) of all problems. That is, skill is defined as a percentile in the probability distribution so that each problem has a difficulty level \(x\) distributed uniformly on \([0, 1]\) and can be solved by its producer if \(x < z\). Let \(F(z)\) and \(f(z)\) denote, respectively, the c.d.f.

\(^2\)See also Ottaviani and Sorensen (2006) and Inderst and Ottaviani (2012).
and the density function of skill in the population. We assume \( f(z) > 0 \) for all \( z \in [0,1] \).

**Occupations.** Individuals sort into two occupations: producers and consultants. Producers generate a problem each, which they attempt to solve. Consultants generate no problems of their own. Instead, consultants provide advice to producers on the problems they failed to solve. Communicating with producers on these problems consumes \( h \) units of the consultant’s time, with \( h \in (0,1) \). Each solved problem generates a value of \$1\ and each unsolved problem is worth \$0\.

Suppose that an producer of skill \( z \) has an unsolved problem and that unsolved problem is allocated to a consultant of type \( m > z \). Then the probability that the consultant can solve the problem is:

\[
\text{Prob}[x < m \mid x > z] = \frac{\int_z^m dx}{\int_z^1 dx} = \frac{m - z}{1 - z}
\]

Define this ex post expected value \( \pi(m, z) = \frac{m - z}{1 - z} \). Note that this function is supermodular (i.e. \( \pi_{m,z} > 0 \)). Intuitively the value of additional talent on a consultant is higher when confronting the problems left unsolved by a more talented producer, which are harder.

From an ex ante standpoint, an producer with skill \( z \) will leave a fraction \((1 - z)\) of his problems unsolved. Thus a single consultant of type \( m \) with a unit of time is able to deal with \( n(z) = \frac{1}{h(1-z)} \) producers of type \( z \). Their ex ante team (conformed of 1 consultant and \( n(z) \) producers) expected output is then:

\[
n(z)z + n(z)(1 - z)\pi(m, z) = \frac{z}{h(1-z)} + \frac{m - z}{h(1 - z)} = \frac{m}{h(1-z)}
\]

Define this ex ante value as \( \Pi(m, z) = \frac{m}{h(1-z)} \). Note that this function is also supermodular, i.e. \( \Pi_{m,z} > 0 \). The key distinction between both is that an producer with a higher talent is *more valuable* from an ex-ante standpoint (\( \Pi_{z} > 0 \)), he is less valuable from an ex post perspective, as unsolved problem by a more talented producer is de facto a harder one, \( \pi_{z} > 0 \).

**Contracts and timing.** We consider two types of contracts, spot, or ex post contracts, and ex ante contracts.

With ex-post contracts agents first choose their occupations, producers then try to solve their problems, those that failed then can enter into a spot contract with a consultant. The hired consultants then try to solve the problems they have. The *spot contract* between a producer an unsolved problem and a consultant has two parts: (1) a fixed fee \( w \in \mathbb{R} \) (a transfer from the producer to the consultant); and (2) a revenue-sharing rule \( \alpha \in [0,1] \) contingent on the consultant solving the problem. That is, the proceeds from the problem are split between the consultant and the producer according to shares \( \alpha \) and \( 1 - \alpha \), respectively. Since there are only two possible outcomes (that the consultant succeeds or fails at solving the problem) \( (w, \alpha) \) contracts are without loss.

With ex-ante contracting, the agents first choose their occupations, then origators and consultants enter into ex-ante contracts, the producer then tries to solve the problem and, if he fails, then the consultant he is partnered with tries to solve the problem. Since *ex-ante* contracts are written
between an producer who is yet to draw a problem and a consultant in addition to \( w \) and \( \alpha \) these contracts also allow for a revenue sharing rule when the producer solves the problem by himself without the help the consultant. In this case the proceeds from the problem are split between the consultant and the producer according to shares \( \gamma \) and \( 1 - \gamma \), respectively with \( \gamma \in [0, 1] \). This richer set of contracts \( \langle w, \gamma, \alpha \rangle \) is without loss when we ex-ante contracts are allowed.

In both the ex-ante or ex-post cases producers can, if the wish, choose not participate in the market for advice and remain independent. We will refer to these agents as "Independents".

Although the main focus of our paper is the case in which the agent’s skill \( z \) is unobservable it is convenient to first analyze the full information benchmark.

### III First best benchmark

#### A Planner’s Problem

Consider a fully informed planner who wishes to maximize social surplus. The planner selects the occupation of each agent and a match between producers and consultants. Without loss, we focus in this section on “ex-ante” matching in which the planner selects a match before producers have attempted their problems, with the understanding that consultants will attempt the unsolved problems of the producers they are matched with. Let \( Z, I \subset [0, 1] \) denote the sets of matched and unmatched producers, respectively, and let \( M \subset [0, 1] \) denote the set of consultants. (Note that the planner wishes to match all consultants because an unmatched consultant would deliver zero output.)

We begin with three observations:

1. **Positive assortative matching.** Given that the output function \( \Pi(m, z) = n(z)m \) is supermodular, the planner’s match must be positive assortative in the sense that if a producer of type \( z \) is allocated to a consultant of type \( m \) it cannot be that a producer of type \( z' > z \) is allocated to consultant of type \( m' < m \). This result is common in the literature and is implication of Theorem 1 in EK (2012). Intuitively, a higher consultant skill is best exploited when applied to a harder problem.

2. **Producer/consultant stratification.** Given that a consultant’s span \( n(z) \) is greater than one regardless of the producer (or producers) \( z \) she is be matched with, the set \( Z \cup I \) of producers must lie below the set \( M \) of consultants.\(^3\) Intuitively, a higher skill is more valuable in the

\(^3\)Formally, the claim is that there is a cutoff \( z_2 \) such that (almost) all types \( m > z_2 \) are consultants and (almost) all types \( z < z_2 \) are producers. Proof: Suppose the claim does not hold – namely, there is a positive measure subset of consultants \( M_0 \) that lies below a positive measure subset of producers \( Z_0 \). Select \( Z_0 \) and \( M_0 \) so that they have equal measures and they each fit in an interval of length \( \varepsilon > 0 \). Denote the measure of these sets \( Q \). Note that no unsolved problem of a type \( z \) in \( Z_0 \) is initially allocated to a consultant with a type \( m \) in \( M_0 \) because that consultant’s time would be wasted. Now (1) make \( Z_0 \) into consultants and allocate to them the unsolved problems originally allocated to the original consultants \( M_0 \) and (2) make \( M_0 \) into originators and allocate their unsolved problems to
hands of a consultant (who applies her skill to $\frac{1}{h}$ problems) than in the hands of a producer (who applies her higher skill to one problem only).

3. *Within-producer stratification.* The set $\mathcal{Z}$ of matched producers must lie below the set $\mathcal{I}$ of unmatched producers.\(^4\) Intuitively, if the planner wishes consultants to attempt only a fraction of the producer’s unsolved problems, it is best that they attempt the problems that are, on average, easiest to solve. These are precisely the problems that come from producers with the lowest types.

From these observations we obtain:

**Lemma 1** The surplus-maximizing allocation consists of a partition of the type space into three intervals, $[0, z_1]$, $(z_1, z_2)$, $[z_2, 1]$, such that:

A. *Agents in $[0, z_1]$ are producers who transfer their unsolved problems to consultants.*

B. *Agents in $(z_1, z_2)$ are “independent” producers who do not transfer their unsolved problems to consultants.*

C. *Agents in $[z_2, 1]$ are consultants.*

D. *There is positive assortative matching between producers in $[0, z_1]$ and consultants in $[z_2, 1]$.*

\[0 \quad \text{Seek Advice} \quad z_1 \quad \text{Independents} \quad z_2 \quad \text{Consultants} \quad 1\]

In order to match producers and consultants the planner selects a pointwise function $M : \mathcal{Z} \rightarrow \mathcal{M}$ that assigns to each matched producer $z$ a consultant $m = M(z)$. Note that positive assortative matching means that $M$ is an increasing function.

\(^4\)Formally, the claim is that there is a cutoff $z_1 \leq z_2$ such that (almost) all $z < z_1$ seek advice and (almost) all $z_1 < z < z_2$ are independent (unmatched) producers.

Proof: Suppose the claim does not hold – namely, there is a positive measure subset of independent producers $Z_0$ that lies below a positive measure subset of matched producers $Z_1$. Select $Z_0$ and $Z_1$ so that they leave the same mass of unsolved problems. Now swap the roles of $Z_0$ and $Z_1$. Since all producers with types in $Z_0$ ask easier questions, the associated consultants solve them more frequently while using the same amount of time, which strictly increases social surplus. QED
Note that the mass of problems left unsolved by producers of type $z$ (namely, $(1-z)f(z)$) need not be equal to the total time available to consultants of type $m$ (namely, $\frac{1}{h}f(m)$). For that reason, a given pointwise match between $z$ and $m$ cannot be interpreted literally. What $M$ does indicate literally is that any given positive measure interval $(z-\varepsilon, z+\varepsilon)$ in $Z$ is matched with the positive measure interval $(M(z-\varepsilon), M(z+\varepsilon))$ in $M$, with the function $M$ satisfying the time budget constraint

$$\int_{z-\varepsilon}^{z+\varepsilon} (1-z)dF(z) = \frac{1}{h} \int_{M(z-\varepsilon)}^{M(z+\varepsilon)} dF(m),$$

where the L.H.S. is the total mass of problems left unsolved by producers in $(z-\varepsilon, z+\varepsilon)$ and the R.H.S. is the total time available to consultants in $(M(z-\varepsilon), M(z+\varepsilon))$. Pointwise matching can then be interpreted as the limiting case where $\varepsilon$ converges to zero.

From remark 1 and the constraint that the total mass of problems left unsolved by producers in the interval $[0, z_1]$ must equal the total time available to consultants in the interval $[z_2, 1]$, we learn that the Planner’s problem reduces to selecting a single variable: the cutoff type $z_1$. Let $M(z; z_1)$ denote the unique increasing matching function that, for a given arbitrary cutoff $z_1$, matches $[0, z_1]$ with $[z_2, 1]$ while satisfying the appropriate time constraint. This function is characterized by the boundary conditions $M(0; z_1) = z_2$, $M(z_1; z_1) = 1$, and the time constraint:

$$\int_{z}^{z_1} (1-z)dF(z) = \frac{1}{h} \int_{M(z; z_1)}^{1} dF(m) \text{ for all } z \leq z_1,$$

where the L.H.S. is the total mass of problems left unsolved by producers in $[z, z_1]$ and the R.H.S. is the total time available to consultants in $[M(z; z_1), 1]$.

Given this set up, we can uniquely characterize the first best as follows:

**Lemma 2** The first-best allocation, uniquely described by cutoffs $z_1^*, z_2^*$, and the increasing matching function $M(z; z_1^*)$, takes one of two forms:

1. $z_1^* < z_2^*$ and

$$z_2^* + \int_{z_2^*}^{1} \frac{1}{h} \frac{1}{1 - Z(m; z_1^*)} dm = \frac{1}{h}.$$

2. $z_1^* = z_2^* = z^*(h)$ and

$$z_2^* + \int_{z_2^*}^{1} \frac{1}{h} \frac{1}{1 - Z(m; z_1^*)} dm \leq \frac{1}{h}.$$

Moreover, there exists a cutoff value $h(F) \in (0, 1)$ such that case 1 (respectively, case 2) arises whenever $h > h(F)$ (respectively, $h \leq h(F)$).
Proof. The planner’s problem is

\[
\max_{z_1} \int_0^{z_1} M(z; z_1) dF(z) + \int_{z_1}^{M(0; z_1)} zdF(z)
\]

s.t.

\[
1 - F(M(z; z_1)) = h \int_z^{z_1} (1 - z) dF(z) \text{ for all } z \leq z_1,
\]

where \( z^*(h) \) uniquely satisfies

\[
1 - F(z^*(h)) = h \int_0^{z^*(h)} (1 - z) dF(z).
\]

The FOC for \( z_1 \) is

\[
\int_0^{z_1} \frac{\partial}{\partial z_1} M(z; z_1) dF(z) + M(0; z_1) f(M(0; z_1)) \frac{\partial}{\partial z_1} M(0; z_1) + (1 - z_1) f(z_1) = \lambda,
\]

and the complementary slackness condition for the second constraint is

\[
\lambda [z^*(h) - z_1] = 0.
\]

Since \( f(M(z; z_1)) \frac{\partial}{\partial z_1} M(z; z_1) = -h (1 - z_1) f(z_1) \) the FOC becomes:

\[
-\int_0^{z_1} h (1 - z_1) f(z_1) \frac{f(z)}{f(M(z; z_1))} dz - M(0; z_1) h (1 - z_1) f(z_1) + (1 - z_1) f(z_1) = \lambda
\]

\[
\Leftrightarrow
\]

\[
-\int_{z_2}^{1} \frac{1}{h} \frac{1}{1 - Z(m; z_1)} dm - M(0; z_1) + \frac{1}{h} = \lambda \frac{1}{h(1 - z_1) f(z_1)}
\]

It follows that, at the optimum, either: (1) \( z_1 < z_2 \) (namely \( z_1 < z^*(h) \)), \( \lambda = 0 \), and

\[
z_2 + \int_{z_2}^{1} \frac{1}{h} \frac{1}{1 - Z(m; z_1)} dm = \frac{1}{h},
\]

or (2) \( z_1 = z_2 \) (namely \( z_1 = z^*(h) \)), \( \lambda \geq 0 \), and

\[
z_2 + \int_{z_2}^{1} \frac{1}{h} \frac{1}{1 - Z(m; z_1)} dm \leq \frac{1}{h}.
\]

The last part of the lemma follows from noting that the marginal value of increasing \( z_1 \) is increasing in \( \frac{1}{h} \).  

Intuitively, as \( z_1 \) increases, the marginal gain on the extensive margin is that some producers are
now being matched with a better solver and having all of their problems solved, while the loss is
the output that independents $z_2$ would have produced on their own. The losses are the sum of the
marginal output loss for every existing producer, who is know matched with a worse solver– they
are all being pushed down to worse matches. Since $(m - z)/(1 - z)$ is the conditional probability
a problem is solved by $m$ given that it was not solved by $z$, $\frac{1}{1 - Z(m; z_2)}$ is the marginal value of the
expert skill for $z$.

It is easy to see from this condition that the share of agents that participate in the consulting
markets increases as communication costs decline. When communication costs are high, agents
operate in autarky, i.e., if they confront a problem, they try to solve it on their own, and if they
cannot, they drop it. As communication costs decrease, the share of agents who seek help with their
problems increases monotonically. Similarly, the share of agents who produce on their own and do
not seek advice (independents) decreases monotonically.

**Corollary 1** The measure of independent agents is increasing in $h$

### B Competitive Equilibrium and One-Sided Informational Asymmetries

When there is no asymmetric information, the first welfare theorem implies the first best can be
attained in a decentralized way as a competitive equilibrium. There are many different decentral-
izations that can implement the first best, and they are all equivalent in the allocation that they
support, which is unique. In particular, we consider two decentralizations that are readily inter-
pretable and will be useful subsequently. They differ in the agent who obtains the residual income
from solving the problem. As a result, they deal with asymmetric information differently.

**A Consulting Market**

**Definition 1** A competitive equilibrium in a consulting market is given by an occupation choice
for all agents, a matching function that determines the choices of consultants by producers and a
type dependent non-contingent price schedule for consulting services $\{w(z)\}$ such that:
1) All agents choose their occupation optimally.
2) There is no excess demand or supply of consulting services.

In a consulting services market, producers hire consultants of skill $z_s$ for a fixed fee $w(z_s)$. pro-
ducers are the residual claimants to output and the expected earnings of a producer of type $z$ who
hires a consultant $z_s$ are:

$$W^c_p(z) = \max_{z_s} F(z) + (1 - F(z)) \left( \frac{F(z_s) - F(z)}{1 - F(z)} - w(z_s) \right)$$

So originatos care about the choice of partners. Note, on the other hand, that the earnings of
consultants do not depend on who they match with; their earnings are simply determined by the
equilibrium consulting fee:

\[ W(z) = \frac{w(z)}{h} \]

The first welfare theorem is enough to show that the competitive equilibrium achieves the first best. Note that in this decentralization, consultants do not make any choices, so they do not need to observe anything. Thus, suppose that the producers’ skill is unobservable, but the skill of consultants is observable. This could be the case, for example, if consultants have developed a reputation that allows agents to know who is knowledgeable and who is not, whereas problem originatos and their quality are unknown. In this case, the consulting market we have just described would still work exactly in the way that we suggested. We state this result in the following remark.

**Remark 1** Under one sided asymmetric information, where consultant skill can be observed but producer skill cannot, the consulting services market still attains the first best.

**A Referral Market**

**Definition 2** A competitive equilibrium in a referral market is given by an occupation choice for all agents, a matching function that determines the choices of problems to buy by consultants and a type dependent non-contingent price schedule for unsolved problems \( r(z) \) such that:

1) All agents choose their occupation optimally.
2) There is no excess demand or supply of unsolved problems.

In a referral market producers transfer the entire residual ownership of the problem to the consultant, in exchange for a fixed referral price \( r(z) \). The earnings of a producer of type \( z \) are given by:

\[ W_p^r(z) = F(z) + (1 - F(z))r(z) \]

Which, in particular, do not depend on the consultant type that buys their unsolved problems. On the other hand, the earnings of consultants of type \( z \) who buy problems from producers of skill \( z_p \) are

\[ W_s^r(z) = \max_{z_p} \frac{1}{h} \left( \frac{F(z) - F(z_p)}{1 - F(z_p)} - r(z_p) \right) \]

Which do depend on both their type and the type they match with. Analogously to the consulting market, the equilibrium in this market does not require observing the skill of consultants. This means that a referral market can achieve the first best when the consultant’s skill is unobservable. For example, suppose all agents can see the skill of agents less skilled than themselves. Then, by making the informed agents, i.e. the consultants, the residual claimants results in the first best.

**Remark 2** Under one sided asymmetric information, where only the skill of producers can be observed (for example, all agents can observe the skill of those less skilled than themselves) the referrals market still attains the first best.
Thus, straightforward institutional arrangements can achieve efficiency if the informational problems are only one sided. In general, in bilateral relationships, letting the party with private information be the residual claimant allows us to achieve efficiency. We have shown that a similar logic extends to this two sided market. As long as the market is set up so that prices are based on the observable type, i.e., a referral market when the producer’s type is observable or a fee based market for advice when the consultant’s type is observable, equilibrium prices will induce the side of the market with private information to self-select the efficient match. This also highlights that even though the intrinsic trade in two markets might be the same, the way they are organized might be very different if the information asymmetries differ across them.

C Surplus distribution

The agents who gain the most from being able to provide and seek advice are those at the extremes, that is the highest skill and lowest skill agents. The most skilled agents benefit from being able to leverage their skill to solve problems for several other agents. The least skilled agents, who on their own can produce very little, benefit most from combining their time with the talent of others. Agents in between, as producers, pose problems that are very hard. As problem solvers, are not very skilled so do not add a lot of value to unsolved problems. Thus ‘quality’ is here decreasing with regard to skill on the advice-seeking side of the market. At the margin, producer $z_1$ is indifferent with regard to seeking advice; to him, advice has no value, i.e., his earnings are the same with or without the advice. On the other side of the market, quality increases with skill, i.e., a better agent can provide better advice and thus benefits most from the market for skills. The following Figure illustrates this result.

Figure 1: Market Allocation (the curved line) versus Autarchy (the straight, 45°line)
IV Two sided adverse selection when the first best prescribes independents

Here we consider our main case of interest in which the surplus-maximizing allocation calls for a positive measure interval of independents \( I = (z_1^*, z_2^*) \). We are interested of characterizing the family, or families, of contracts \( \langle w(z), \gamma(z), \alpha(z) \rangle_{z \in Z} \) (of which ex post contracts are a special case) that implement the first best.

For notational simplicity, denote the first best matching function by \( M(z) = M(z; z_1^*) \) for all producer types \( z \in Z \) and denote the inverse of this function by \( Z(m) = M^{-1}(m) \) for all consultant types \( m \in M \). Namely, in the first best, a producer of type \( z \) is matched with a consultant of type \( m = M(z) \) and a consultant of type \( m \) is matched with a producer of type \( z = Z(m) \).

A Preliminaries

The ex-ante expected payoff for a producer of type \( z \) who selects contract \( \langle w(z'), \gamma(z'), \alpha(z') \rangle \) is given by:
\[
R(z, z') = w(z') + z(1 - \gamma(z')) + (M(z') - z)(1 - \alpha(z')).
\]
The ex-ante expected payoff for a consultant of type \( m \) who selects contract \( \langle w(z'), \gamma(z'), \alpha(z') \rangle \) is given by:
\[
S(m; Z(m')) = \frac{1}{h[1-z']} \left[ -w(z') + z'\gamma(z') + (m - z')\alpha(z') \right] \text{ for } z' = Z(m').
\]
With slight abuse of notation, define \( R(z) = R(z, z') \) and \( S(m) = S(m, Z(m)) \), which are the payoffs of a producer and a consultant, respectively, who report their types truthfully.

Note that, at the first best, team surplus must be fully allocated within each team:
\[
R(z) + h[1-z]S(M(z)) = M(z). \quad (TS)
\]
The following are necessary and sufficient condition for a family of ex-ante contracts \( \langle w(z), \gamma(z), \alpha(z) \rangle_{z \in Z} \) to implement the first best:

First, the relevant incentive constraints within the optimal interval or producers must be met:
\[
R(z) \geq R(z, z') \text{ for all } z, z' \in Z. \quad (IC_1)
\]
Second, the relevant incentive constraints within the optimal interval of consultants must be met:
\[
S(m) = S(m, Z(m')) \text{ for all } m, m' \in M. \quad (IC_2)
\]
Third, every individual must find it optimal to select the first best occupation:

\[
R(z) \geq \max \left\{ z, \max_{m'} S(z, Z(m')) \right\} \quad \text{for all } z \in Z, \quad (i)
\]

\[
S(m) \geq \max \left\{ m, \max_{z'} R(m, z') \right\} \quad \text{for all } m \in M, \quad (ii)
\]

\[
z \geq \max \left\{ \max_{z'} R(z, z'), \max_{m'} S(z, Z(m')) \right\} \quad \text{for all } z \in Z. \quad (iii)
\]

We refer to these three inequalities collectively as the occupational sorting constraints.

**Remark 3** The family of contracts \( \langle w(z), \gamma(z), \alpha(z) \rangle_{z \in Z} \) satisfies the producer incentive constraints \((IC_1)\) if and only if

\[
R(z) = z^*_2(1 - h) + \int_0^z [\alpha(t) - \gamma(t)] \, dt \quad \text{for all } z \in Z, \quad (E_R)
\]

\[
\alpha(z) - \gamma(z) \text{ is nondecreasing in } z. \quad (M_R)
\]

**Proof.** Omitted as it a straightforward implication of the Envelope Theorem (Milgrom and Segal, 2002). ■

**Remark 4** The family of contracts \( \langle w(z), \gamma(z), \alpha(z) \rangle_{z \in Z} \) satisfies the consultant incentive constraints \((IC_2)\) if and only if

\[
S(m) = z^*_2 + \frac{1}{h} \int_{z^*_2}^m \frac{\alpha(Z(t))}{1 - Z(t)} \, dt \quad \text{for all } m \in M, \quad (E_S)
\]

\[
\frac{\alpha(z)}{1 - z} \text{ is nondecreasing in } z. \quad (M_S)
\]

**Proof.** Omitted as it a straightforward implication of the Envelope Theorem (Milgrom and Segal, 2002). ■

**Lemma 3** Suppose the family of contracts \( \langle w(z), \gamma(z), \alpha(z) \rangle_{z \in Z} \) satisfies \((IC_1)\) and \((IC_2)\). Then, that family of contracts satisfies the occupational sorting constraints \((i)-(iii)\) if and only if

1. The equilibrium payoff functions are continuous at \( z^*_1 \) and \( z^*_2 \):

\[
R(z^*_1) = z^*_1, \quad S(z^*_2) = z^*_2. \quad (PC)
\]

2. The directional derivatives of the equilibrium payoff functions at \( z^*_1 \) and \( z^*_2 \) satisfy the Envelope Theorem:

\[
R'(z^*_1-) \leq 1 \quad (DD_R)
\]

\[
S'(z^*_2+) \geq 1 \quad (DD_S)
\]
Proof. Necessity ($\rightarrow$). On the one hand, (PC) follows from the fact that, regardless of their occupational choice, the individuals’ equilibrium payoffs must be continuous in type. On the other hand, (DD$_R$) and (DD$_S$) follow directly from theorem 1 in Milgrom and Segal (2002): namely, at any given type, the right hand derivative of the equilibrium payoff function (which is an upper envelope) must be no smaller than its left hand derivative.

Sufficiency ($\leftarrow$). Since payoffs are single crossing in type, it suffices that occupational sorting constraints are met at the occupational boundaries $z^*_1$ and $z^*_2$, which is turn is guaranteed by (PC).

We are now ready to prove a result that is a central building block for the results that follow:

Lemma 4 Suppose the family of contracts $(w(z), \gamma(z), \alpha(z))_{z \in Z}$ implements the first best allocation. Then, $\alpha(z) = 1$ for all $z \in Z$.

Proof. One the one hand, the planner’s FOC implies that

$$
\frac{1}{\tilde{h}} - z^*_2 = \int_{z^*_2}^{1} \frac{1}{\tilde{h} 1 - Z(t)} dt.
$$

(a)

On the other hand, the producer envelope condition (E$_S$) implies that

$$
S(1) - z^*_2 = \int_{z^*_2}^{1} \frac{1}{\tilde{h} 1 - Z(t)} \alpha(Z(t)) dt.
$$

(b)

From (TS) and the fact that $R(z^*_1) = z^*_1$ it follows that $S(1) = \frac{1}{\tilde{h}}$. Consequently, given that $\alpha(z) \in [0, 1]$ for all $z$, equitations (a) and (b) can only be simultaneously satisfied if $\alpha(z) = 1$ for all $z$.

B Ex-post (spot) contracting

Recall that ex-post contracting is a special case of ex-ante contracting in which $\gamma(z)$ is restricted to equal $w(z)$ for all $z \in Z$. The following result shows that spot contracting can never achieve the first best when such allocation calls for independent producers:

Proposition 1 Suppose the first-best allocation prescribes an interval of independent producers. The first best allocation cannot be supported by ex-post contracting.

Proof. Recall that (PC) requires that $S(z^*_2) = z^*_2$ and Lemma 4 implies that $\alpha(0) = 1$. It follows that $w(0) > 0$ and so every type $z \in I$ would gain by pretending to be of type 0 and thus selling her unsolved problems to a consultant of type $m = M(0) = z^*_2$.

C Ex-ante contracting

This section presents the main result of our paper. Define $\beta(z) = 1 - \gamma(z)$, which we interpret as a bonus paid by each consultant of type $M(z)$ to each producer of type $z$ he is matched with per
Problem solved by the producer.

**Theorem 1** Suppose the first-best allocation prescribes an interval of independent producers. The following family of contracts uniquely implements the first best allocation: for all $z \in \mathcal{Z}$,

$$
\begin{align*}
\alpha(z) &= 1, \\
\beta(z) &= hS(M(z)), \\
w(z) &= M(z) - hS(M(z)),
\end{align*}
$$

where, for all $m \in [z_2, 1]$,

$$
S(m) = z_2 + \frac{1}{h} \int_{z_2}^{m} \frac{1}{1 - Z(m)} dm.
$$

**Proof.** It follows from Lemma 4 that $\alpha(z) \equiv 1$ uniquely satisfies $(PC_S)$, $(DD_S)$, $(E_S)$, and $(M_S)$. Now differentiate $(TS)$ to obtain

$$
R'(z) = M'(z) - h [1 - z] S'(M(z)) M'(z) + h S(M(z))
= M'(z) [1 - \alpha(z)] + h S(M(z)),
$$

and combine the above equality with $(E_R)$ to obtain

$$
\beta(z) = 1 - \gamma(z) = h S(M(z)). \quad (\alpha - \gamma)
$$

Note that $PC_R$ is satisfied by construction, $(M_R)$ is satisfied because $S$ is an increasing function, and $(DD_R)$ is satisfied because $\alpha(z_1) - \gamma(z_1) = 1$. ■
Case B: The first best allocation prescribes no independents

Here we consider the simpler case in which the surplus-maximizing allocation calls for no independents. As before, we are interested in characterizing the family, or families, of ex-ante contracts \((w(z), \gamma(z), \alpha(z))_{z \in Z}\) (of which ex-post contracts are a special case) that implement the first best.

D Preliminaries

The following are necessary and sufficient conditions for a family of ex-ante contracts \((w(z), \gamma(z), \alpha(z))_{z \in Z}\) to implement the first best:

First, as before, the incentive constraints \((IC_1)\) within the optimal interval or producers must be met. Second, the \((IC_2)\) incentive constraints within the optimal interval of consultants must be met. Third, every individual must find it optimal to select the first best occupation:

\[
R(z) \geq \max_{m'} S(z, Z(m')) \text{ for all } z \in Z, \quad (i)
\]
\[
S(m) \geq \max_{z'} R(m, z') \text{ for all } m \in M. \quad (ii)
\]

We refer to these two inequalities collectively as the occupational sorting constraints. Fourth, the equilibrium payoff of each type must be no smaller than the autarky payoff for that type:

\[
R(z) \geq z \text{ for all } z \in Z, \quad S(m) \geq m \in M. \quad (IR)
\]

Lemma 5 Suppose the first best prescribes no independents \((I = \emptyset)\) and suppose the family of contracts \((w(z), \gamma(z), \alpha(z))_{z \in Z}\) satisfies \((IC_1)\) and \((IC_2)\). Then, that family of contracts satisfies the occupational sorting constraints \((i)\) – \((ii)\) if and only if

1. The equilibrium payoff functions are continuous at \(z_1^*\):

\[
R(z_1^*) = S(z_1^*). \quad (PC)
\]

2. The directional derivatives of the equilibrium payoff functions at \(z_1^*\) satisfy the Envelope Theorem:

\[
R'(z_1^*-) \leq S'(z_1^*+). \quad (DD)
\]

Proof. Necessity \((\Rightarrow)\). First, \((PC)\) follows from the fact that the individuals’ equilibrium payoffs must be continuous in their type. Second, \((DD)\) follows directly from theorem 1 in Milgrom and Segal (2002).

Sufficiency \((\Leftarrow)\). Since payoffs are single crossing in type, it suffices that occupational sorting constraints are met at the occupational boundary \(z_1^*\), which is in turn guaranteed by \((PC)\).
E Ex-post (spot) contracting

When the first best does not call for independent producers, ex-post contracting may or may not achieve the first best, depending on the distribution of types $F$ and the value of $h$. In general, it is difficult to very analytically whether the first best can be achieved for a given set of parameters.

Example 1 Suppose the first-best allocation prescribes no independent producers. When types are distributed uniformly, the first best allocation can be supported by ex-post contracting.

F Ex-ante contracting

This section shows that ex-ante contracting always supports the first best. The difference with respect to the case in which the first best prescribed independent producers is that the first best may be in principle be supported by more than one family of contracts.

Theorem 2 Suppose the first-best allocation prescribes no independent producers ($\mathcal{I} = \emptyset$). The following family of contracts implements the first best allocation: for all $z \in \mathcal{Z}$,

\[
\alpha(z) = 1, \\
\beta(z) = hS(M(z)), \\
w(z) = M(z) - hS(M(z)),
\]

where, for all $m \in \mathcal{M}$,

\[
S(m) = R(z^*_1) + \frac{1}{h} \int_{z^*_1}^{m} \frac{1}{1 - Z(m)} dm.
\]

Proof. Let $z^* = z^*_1$ and set $\alpha(z) = 1$ for all $z$. $S(1)$ and $R(z^*)$ must satisfy two conditions. First, the total team surplus condition

\[
R(z^*) + h(1 - z^*) S(1) = 1
\]

Second, the envelope condition

\[
S(1) = R(z^*_1) + \frac{1}{h} \int_{z^*_1}^{1} \frac{1}{1 - Z(m)} dm
\]

Combining both equations delivers the unique candidate value for $R(z^*)$, namely,

\[
\frac{1 - R(z^*)}{h(1 - z^*)} = R(z^*_1) + \frac{1}{h} \int_{z^*_1}^{1} \frac{1}{1 - Z(m)} dm
\]

(+) Two conditions must be met:

1. Directional-derivative condition at $z^*$: $R'(z^*) = 1 - \gamma(z^*) = hS(1) \leq \frac{1}{h} \alpha(0) = \frac{1}{h}$. 

2. Participation for all types, which boils down to: (a) \( R(z^*) \geq z^* \); (b) \( R'(z) = 1 - \gamma(z^*) = hS(z) \leq 1 \) for all \( z \).

Note that \( R(z^*) \geq z^* \) implies that \( S(1) \leq \frac{1}{h} \) and so \( hS(1) \leq 1 \), which implies that conditions 1 and 2(b) are redundant.

Since the L.H.S. of (*) is decreasing in \( R(z^*) \) and the R.H.S. is increasing in \( R(z^*) \), the remaining condition \( R(z^*) \geq z^* \) is met whenever

\[
\frac{1 - z^*}{h(1 - z^*)} \geq z^* + \frac{1}{h} \int_{z^*}^{1} \frac{1}{1 - Z(m)} \, dm
\]

\( \iff \)

\[
\frac{1}{h} \geq z^* + \frac{1}{h} \int_{z^*}^{1} \frac{1}{1 - Z(m)} \, dm
\]

which is precisely the inequality established above for the first best (case 2). Implications and Evidence

V Conclusion

Advice or consulting markets represent an increasing share of economic activity.\(^5\) These markets entail the separation of production activities from (part of the) specialized knowledge required to produce and sell goods. In this paper, we study the ability of these markets to operate efficiency. The problem they face is that these markets operate under double sided adverse selection: those asking for advice cannot observe the true quality of the consultants offering help, while those offering advice cannot observe the difficulty of the questions posed. What is special about these markets is that the lemons are those in the middle— their talent is too low to be consultants but their problems are too hard for them to participate by asking for help.

We have shown that when efficiency calls for some self-employed agents, spot contracting is inefficient. Instead, firm-like contracts implement an efficient allocation of talent – uniquely so when efficiency calls for self-employed agents. The optimal firm-like contract induces a hierarchy in which the head (the consultant) is the full residual claimant of all problems, and pays a fixed wage plus bonus to producers that is increasing in their talent. That is, incentives become higher powered as employee talent increases.

The significance of the bonus is that it deters independent employees from entering the market. Essentially, the independents would like to imitate the worst producers (by playing dumb) – but

\(^5\) Over the last fifty years, the share of the service sector in the GDP of Western economies has grown substantially, e.g. in the USA by about 20 points. Approximately half of this increase is the result of the increase in the share of professional and business services such as computer services, consulting and legal services (Herrendorf, Rogerson and Valentinyi, 2009). Fuchs (1965, p. 3) argued against demand shifts as the reason for the transformation and contended that ‘growth in intermediate demand for services by goods producing industries’ would account for 10% of the shift. Herrendorf, Rogerson and Valentinyi (2009), argue that outsourcing is unlikely to explain the entire trend, as business services only account for half of the increase in the expenditure share of services.
this would involve taxing them when they can in fact solve the problem. When a consultant taxes a producer for the problems that this producer manages to solve, producers with high types (those who would optimally remain outside of the market) have a new incentive to stay away.

This simple characterization of the optimal ex ante contract provides a rationale for the organization of firms and the structure of compensation in the knowledge intensive sector (law firms, consulting firms, advertising agencies, private equity firms, etc.). Like in our paper, firms in that sector are structured as hierarchies, with fewer partners than associates. Partners retain the entire equity in the firm and earn the residual income, and pay increasing bonus depending on output to the associates. According to our analysis, this is the unique contracting structure that attains first best in the presence of double sided informational asymmetries.
REFERENCES


