Optimal Contracting and the Organization of Knowledge*

William Fuchs
Berkeley-Haas and IZA

Luis Garicano
LSE and CEPR

Luis Rayo
LSE and CEPR

January 27, 2014

Abstract

We study contractual arrangements that support efficient production in a knowledge-intensive economy. Such production is plagued by informational problems, since both the difficulty of the questions posed to experts and the knowledge of those experts are hard to assess. We show that spot markets are, in general, not efficient since lemons (in this market, self-employed agents with intermediate knowledge) cannot be appropriately excluded. However, an ex-ante, firm-like contractual arrangement, involving hierarchies in which experts are full residual claimants of output and compensate non-experts via incentive contracts, is guaranteed to deliver the first best (uniquely so whenever some agents are self-employed). This simple characterization of the optimal ex-ante arrangement suggests a rationale for the organization of firms and the structure of compensation in knowledge-intensive sectors.

*We are grateful to Marina Halac, David Rahman, Esteban Rossi-Hansberg, John Sutton, our editor Marco Ottaviani, and three anonymous referees for their valuable suggestions. We also thank seminar participants at the AEA meetings, Berkeley, Boston University, Carlos III, CEMFI, Chicago Booth, Naples, Frankfurt, University of Washington, Washington University, and the Utah WBEC for their comments. Miguel Espinosa provided valuable research assistance. Garicano would like to thank the Toulouse Network in Information Technology for financial support. A previous version of this paper was circulated under the title "Trading Know-How."

1 Introduction

A key role of markets and other organizations is ensuring that society uses the knowledge available to economic agents optimally (Hayek, 1945). Doing so requires conserving the superior knowledge of experts for the right questions; that is, efficiently matching questions and expertise. A solution to this problem is a "knowledge-based hierarchy" in which experts lever their knowledge by having less expensive workers deal with routine tasks. Firms and other organizations often rely on such hierarchies by placing agents in different positions according to their expertise.[1]

The allocation of knowledge is complicated by the presence of two-sided information asymmetries: both the difficulty of the problems posed to experts and the knowledge of those experts may be hard to assess. As a result the market may be unable to ensure an efficient use of expert time. In this paper, we study whether or not there exist contractual arrangements that support an efficient allocation despite these informational asymmetries, and, if so, the form that such arrangements take.

We carry out our analysis in a general-equilibrium setting, in which agents, each endowed with a privately-observed knowledge level and a unit of time, choose their occupation (production or consulting) and whether or not 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 faces a problem that is too hard relative to her knowledge level, she may seek help from a consultant.

We start by deriving, as a benchmark, the first-best allocation under full information. This allocation has the following features. First, agents are split into three sets: *matched producers* (who receive advice from consultants upon failing to solve their problems); *self-employed producers* (who never receive advice); and *consultants*. Second, consultants are more knowledgeable than both types of producers, and self-employed producers, in turn, are more knowledgeable than matched producers. Accordingly, self-employed producers have intermediate knowledge levels. Third, matching between producers and consultants is positive assortative: more knowledgeable consultants devote their time to problems that are expected to be more difficult, namely, those problems that the most knowledgeable producers failed to solve on their own. Fourth, whenever communication between consultants

[1] Garicano (2000) studies knowledge-based hierarchies in which ex-ante homogeneous workers acquire knowledge and choose their positions in the hierarchy. Garicano and Rossi-Hansberg (2006) analyze the problem of knowledge acquisition and optimal organizational structure with heterogeneous agents. Both these papers focus on economies with full information.

and producers is sufficiently inexpensive in terms of the time it consumes, the set of self-employed producers is empty. In this case, even the hardest problems are worth a second try and all agents participate in the market.

Intuitively, team production is most effective when the most knowledgeable agents (who specialize in consulting on difficult problems) are matched with the least knowledgeable ones (who specialize in production and get help from consultants only when needed). Agents with intermediate knowledge, in contrast, contribute to team production the least. The existence of self-employed producers, as we shall see, is the key obstacle to an efficient spot market. It is also one of the features that distinguishes the present setting from a standard asymmetric-information setting: in our model, the 'lemons' who threaten market efficiency are precisely the agents with intermediate knowledge levels.

Our main focus is on contracting with asymmetric information. We consider two forms of contracts: ex-post (or spot) and ex-ante. Under ex-post contracting, producers seek help from consultants only after facing a problem they cannot solve. Such contracts rely on two pieces of information to condition payments between agents: (i) participation in the contract; and (ii) whether or not the producer solves her problem upon receiving help from the consultant. Under ex-ante contracting, in contrast, consultants and producers match before producers attempt their problems (in a similar way, for example, as contracting inside a firm). Crucially, ex-ante contracts use an extra piece of information to condition payments: whether or not a producer solves her problem on her own.

We find that spot markets (i.e. those with ex-post contracting) are, in general, not efficient. The reason is that self-employed producers may want to overstate their knowledge and become consultants and/or understate their knowledge and seek advice. Excluding them from the market would require that asking for help is costly (to prevent self-employed producers from pretending to have easy unsolved problems) and, at the same time, would require that consultants pay producers up front for the right to offer advice (to prevent self-employed producers from becoming consultants). We show that spot contracts cannot achieve both goals simultaneously. Accordingly, when the first best calls for self-employed producers it cannot be supported by such contracts.[2]

Our main result is that, in contrast to spot markets, ex-ante contracting always supports the first-best allocation, despite agents contracting under two-sided information asymmetries.

[2] This observation may help us understand why recent advances in communication technology have led to the appearance of new types of spot exchanges for problem solving and advice. We return to this point in Section 4.2 below.

Moreover, we show that there is a unique ex-ante arrangement that guarantees an efficient allocation, regardless of the underlying parameters. This arrangement has the following intuitive structure:

1. Each consultant matches, up front, with a number of producers and enters into a bilateral contract with each. Consultants with higher knowledge match with more knowledgeable producers (and with a greater number of producers).

2. Consultants pay each producer in their team a positive fixed fee (e.g. a wage) plus a performance bonus in the event that the producer succeeds at solving her problem on her own. The fixed fee is decreasing, and the bonus is increasing, in the producer's knowledge level. In other words, as producer knowledge grows, incentives become higher-powered. It is this feature of the arrangement that screens producers.

3. Net of the above payments to producers, each consultant is the full residual claimant of all output produced by the team. The total expected payment from the consultant to each producer is increasing in the producer's knowledge at a rate equal to the marginal contribution of producer knowledge to expected team output. It is this feature that screens consultants: they simply "hire" producers at their marginal value.

Crucially, these contracts are able to simultaneously screen agents on both sides of the market. When serving as residual claimants, consultants care only about their total expected payment to producers. Producers, in contrast, care also about the way in which the total expected payment is split between fixed and bonus payments.

These contracts also guarantee that the most knowledgeable (self-employed) producers stay away from the market. On the one hand, those producers are willing to remain self-employed because being matched with a consultant (under the pretense of having lower knowledge) commits them to share their output with the consultant when succeeding on their own. Since such output sharing is, in expectation, more costly for producers with higher knowledge (who have a higher likelihood of success), it precisely induces producers with the highest knowledge levels to remain unmatched. On the other hand, self-employed producers do not wish to become consultants because, as such, they would become the full residual claimants of the unsolved problems they consult on, a feature they dislike because they are less capable, compared to existing consultants, of solving such problems. Crucially, unlike ex-ante contracts, spot contracts do not allow for the above output-sharing commitments that prevent self-employed producers from entering the market.

We show, moreover, that whenever the first best calls for self-employed producers (i.e. consultant time is sufficiently scarce), the above ex-ante contracts are the *only* ones that implement the first best. The reason is that granting full residual claim to consultants maximally deters self-employed producers from wanting to become consultants, as it would force them to internalize the full marginal value of their (relatively-low) knowledge. Such maximal deterrence is in turn required to protect the scarce talent of the most knowledgeable consultants.

Our results are reminiscent of Alchian and Demsetz's (1972) seminal work. While they motivate the appearance of a firm on moral-hazard grounds by arguing that the "specialization in monitoring plus reliance on residual claimant status will reduce shirking" (p. 782), our paper instead shows that specialization in advising and residual-claimant status for consultants solves a double-sided adverse selection problem. Such contracts ensure that experts seek to "employ" producers of the right level (as more knowledgeable producers receive, in expectation, a higher compensation) and, simultaneously, ensure that producers wish to "work" for the right experts (as the power of their incentive pay increases with their reported knowledge level).

Lastly, we consider two extensions. First, we consider the case in which all matching between producers and consultants is controlled by a monopolistic intermediary. We show that the intermediary finds it in her interest to implement the first-best allocation while keeping all rents for herself. This result illustrates an alternative (stylized) resolution to the failure of spot markets. Second, to expose the critical role of contingent payments, we consider the case in which consultants cannot be paid success-contingent fees, but we otherwise allow for general random allocations. We show that, in such case, the market for knowledge breaks down completely, regardless of the potential gains from trade.

No previous study has, to our knowledge, examined the problem of matching consultants and producers under double-sided asymmetric information. The previous literature on consulting markets emphasizes moral hazard in the provision of consultant services. Demski and Sappington (1987) examine the trade-off between productive effort and information-gathering incentives faced by consultants. Wolinsky (1993) considers the type of incentives that must be provided for consultants to recommend a proper treatment, i.e. a minor treatment for minor problems and a major treatment for major problems. Wolinsky shows that specialization is optimal in such a setting. Similarly, Pesendorfer and Wolinsky (2003) study the provision of adequate diagnostic effort by consultants.[3] We depart from this literature in that we abstract from moral hazard altogether and focus instead on double-sided adverse selection.[4]

Garicano and Santos (2004) study a referral problem with one-sided adverse selection, where each production opportunity is always dealt with by a single individual. Agents' skills are known, occupational choices are immaterial, and the agent who first receives an opportunity automatically learns its value. The efficient allocation involves low-skill agents tackling low-value opportunities and high-skill agents tackling high-value ones. Their model also features moral hazard and, indeed, their focus is on the trade-off between moral hazard and efficient problem allocation. In contrast, we study knowledge hierarchies, with two-sided adverse selection, where more than one agent could be involved in solving a particular problem ("management by exception"). Agents' skills are unknown and they can choose their occupation. Moreover, the hierarchy structure exacerbates adverse selection as only the agents with problems that turned out to be hardest are in need of advice from high-skill agents. Finally, in Garicano and Santos (2004) the first best need not be achieved, whereas we show that a simple, firm-like contract always attains the first best.

Our paper is also related to the literature on trade in markets with bilateral asymmetric information. Most of this literature stems from Myerson and Satterthwaite's (1983) seminal analysis of trade between buyers and sellers with private valuations (see also, for example, Lu and Robert, 2001). In this literature, in contrast to our setting, buyers and sellers do not care about each other's types directly, players rely only on non-contingent payments, and matching considerations are not studied. The only prior paper we are aware of that studies equilibria in matching markets with two-sided adverse selection is Gale (2001). There are several important differences between our models. First, Gale takes as exogenous the side of the market agents belong to, while in our model agents select into different occupations. Second, in Gale's paper all agents have equal outside options, while in our setting higher-quality agents have more attractive outside options, which further exacerbates the adverse selection problem. Third, Gale focuses on obtaining refinements that guarantee a fully-separating equilibrium. In contrast, we focus on characterizing contracts supporting the first best.

[3] Taylor (1995) studies how insurance can solve informational asymmetries in a context where only the consultant can determine the necessary treatment.

[4] See also Ottaviani and Sorensen (2006), who study cheap talk by an expert who cares about his reputation, and Inderst and Ottaviani (2012), who study incentives in the form of commissions and kickbacks for advisors who act as intermediaries between consumers and firms.

The paper is also related to the literature on management-worker sorting under full information, and, in particular, to Garicano and Rossi-Hansberg (2004, 2006). These models have been generalized by Eeckhout and Kircher (2012), who study general conditions under which interactions in production between worker and managerial skills generate skill-scale effects and positive sorting under full information. As in these papers, the economy-wide problem we study here is one of matching talent with problems. In contrast to these papers, introducing informational asymmetries and optimal contracting allows us to study the contractual structure underpinning hierarchical team formation and, indeed, leads us to the finding that “firm-like”contracts are the unique ones that allow for such teams to be (optimally) formed under double-sided informational asymmetries. Our analysis is thus novel in its ability to characterize optimal contracting, which was immaterial in the settings above. We also offer a more general characterization of the first best with exogenous knowledge than previous work.

In independent and complementary work, Acemoglu, Mostagir, and Ozdaglar (2014) consider a problem of crowdsourcing. In their model, a principal who owns a set of unsolved problems contracts with agents (either his employees or external contractors) who attempt to solve these problems. They show that the first-best allocation of talent, which involves an endogenous hierarchy of problem solvers, is implementable regardless of informational asymmetries. Their approach differs from our own in two key respects. First, our technologies differ. In Acemoglu et al., in accordance to their focus on crowdsourcing, there is a limited supply of problems and agents cannot advise each other (instead they pass unsolved problems to one another). In our setting, in accordance to our focus on knowledge hierarchies with specialized advisors, the supply of problems is unlimited (i.e. all agents can work on their own problems if they so desire) and agents can communicate with each other, allowing the most knowledgeable among them to withdraw from direct production and instead specialize on advising others. Second, while in Acemoglu et al. the principal who owns all problems is exogenous, we obtain an endogenous principal who acts as an advisor and is optimally assigned the residual claim of output.

The remainder of the paper is organized as follows. Section 2 presents the model setup. Section 3 derives the first-best benchmark and describes full-information competitive equilibria that support it. Section 4, which contains our main results, addresses contracting under asymmetric information. Section 5 presents our two extensions. Finally, Section 6 concludes.

2 Model setup

We model an economy in which, in the spirit of Garicano (2000) and Garicano and Rossi-Hansberg (2006), agents use time and knowledge to solve problems.

Agents, knowledge, and problems. The economy is formed by a continuum of risk-neutral agents, each endowed with one unit of time and an exogenous, privately-known knowledge level $z \in [0,1]$. Knowledge levels are distributed according to a c.d.f. F with associated density f. We assume $f(z) > 0$ for all z. Each agent may or may not engage in production, which requires time and knowledge as inputs. When an agent engages in production, she devotes her full unit of time, and applies her knowledge, to a single problem of difficulty x. The difficulty of each problem is unknown and is distributed uniformly on $[0,1]$, independently across problems.[5] The agent manages to solve the problem on her own if her knowledge exceeds the problem's difficulty level, i.e. $z > x$.

Occupations. Agents sort into one of two occupations: production and consulting. Producers initially attempt to solve their respective problems on their own. Those who fail may opt to seek advice from consultants. Consultants do not spend any time on production. Instead, they specialize on transferring knowledge to producers who seek advice. Transferring knowledge consumes h ("help") units of consultant time per problem, with $h \in (0,1)$, and, for simplicity, consumes no additional producer time. A producer who seeks advice from a consultant manages to solve her problem if the knowledge of the consultant exceeds the problem's difficulty level.

Teams and output. Producers may either never seek advice or form teams with consultants. Teams, in turn, may be formed either ex-ante (before producers attempt their problems) or ex-post (after producers have failed their problems).[6]

Throughout, we describe teams as pointwise matches between a number of producers of a given type z and a consultant of a given type m, while allowing the number of producers in the team to be non integer. The literal interpretation of such pointwise matching is that a small interval of producers $(z - \varepsilon_1, z + \varepsilon_1)$ is matched with a small interval of consultants $(m - \varepsilon_2, m + \varepsilon_2)$ with the requirement that the combined mass of consultants in the team (namely, $F(m+\varepsilon_2) - F(m-\varepsilon_2)$) is h times as large as the mass of problems left unsolved by

[5] Since the distribution F is arbitrary, the assumption that x is distributed uniformly is without loss (i.e. is merely a normalization) under suitable regularity conditions (namely, that both z and x are distributed on $[0,1]$ according to a density function that is everywhere positive).

[6] We restrict attention to teams with a single layer of consultants. See, for example, Garicano and Rossi-Hansberg (2006) for an analysis of multi-layered hierarchies in a full-information setting.

the producers in the team (namely, $h\int_{z-\varepsilon_1}^{z+\varepsilon_1}(1-t)dF(t)$). This requirement ensures, assuming an exact law of large numbers, that the combined measure of consultant time is fully occupied and all producers with unsolved problems receive advice.[7] Pointwise matching (potentially involving a non-integer number of producers) can then be viewed as the limiting case where ε_1 and ε_2 converge to zero.

Note that we restrict from the outset to matching arrangements in which all producers matched with a given consultant have the same type. This restriction is without loss because we seek efficient contractual arrangements, all of which, as we show in Section 3, exhibit (strict) positive-assortative matching.

A solved problem generates a value of \$1 and an unsolved problem generates a value of \$0. Accordingly, the output generated by the various possible arrangements above is as follows:

1. A producer who never seeks advice generates an expected output equal to her knowledge z.

2. From an ex-post standpoint, if a consultant of type m devotes her unit of time to advise $\frac{1}{h}$ producers of type $z < m$, the expected output of such team is

$$\pi(m,z) \equiv \frac{1}{h} Prob[x < m \mid x \geq z] = \frac{1}{h} \cdot \frac{m-z}{1-z}.$$

   We refer to $\pi(m,z)$ as the "ex-post production function." Note that this function is supermodular (i.e. $\pi_{mz} > 0$). Intuitively, a higher consultant knowledge is more valuable, in the margin, when applied to a harder problem. Harder problems, in turn, are precisely those left unsolved by more knowledgeable producers.

3. From an ex-ante standpoint, each producer with knowledge z will need help with probability $(1-z)$. Thus, a consultant of type m is able to form a team with $n(z) \equiv \frac{1}{h(1-z)}$ producers of type z, as, in expectation, these producers need help on $\frac{1}{h}$ problems. (While the number of problems left unsolved by those $n(z)$ producers is random, the fact that matching actually occurs between small intervals of types on either side allows

[7] Note that this interpretation of team formation is meaningful despite the facts that: (1) the number of problems left unsolved by a single producer is not deterministic; (2) the number of problems that a single consultant can advise on (namely, $\frac{1}{h}$) need not be an integer; and (3) the density of problems left unsolved by producers of type z (namely, $(1-z)f(z)$) is generically not equal to the density of problems that consultants of type m can advise on (namely, $\frac{1}{h}f(m)$).

us to treat such number of unsolved problems as deterministic.) The expected output of such team is

$$\Pi(m,z) \equiv n(z)z + n(z)(1-z)\frac{m-z}{1-z} = n(z)m.$$

We refer to $\Pi(m,z)$ as the "ex-ante production function." Note that this function is also supermodular (i.e. $\Pi_{mz} > 0$). Intuitively, since the consultant's type only matters when advice is given, Π inherits the supermodularity of π.

Note a key difference between the ex-ante and ex-post production functions. A producer with a higher type is more valuable from an ex-ante standpoint ($\Pi_z > 0$), as she is less likely to require advice. In contrast, a producer with a higher type is less valuable from an ex-post standpoint ($\pi_z < 0$), as a problem left unsolved by a more knowledgeable producer is, in expectation, more difficult.

Contracts and timing. Contracts are pairwise agreements between a consultant and each producer in her team. We consider two types of contracts, which differ in their timing and thus in the information that can be used to condition payments: ex-post (or spot) and ex-ante.

Under ex-post contracting, timing is as follows. First, agents choose their occupations and producers attempt their problems. Second, each producer who failed to solve her problem decides whether or not to enter into a spot contract with a consultant. If she does not, the producer's problem remains unsolved. If she does, the producer receives advice and a fixed transfer $w \in \mathbb{R}$ from the consultant. In addition, the proceeds from the problem are split between the consultant and the producer according to shares α and $1-\alpha$, with $\alpha \in [0,1]$. Since there are only two possible outcomes (the problem is either solved or not) an ex-post contract of the form $\langle w, \alpha \rangle$ is without loss.[8]

Under ex-ante contracting, timing is as follows. First, agents choose their occupations. Second, each producer may or may not enter into an ex-ante contract with a consultant. If she does not, she remains in autarchy. If she does, the producer receives a fixed transfer $\omega \in \mathbb{R}$ up front from the consultant. Third, each producer attempts her problem and seeks advice if needed. If the producer solves the problem on her own, the producer obtains a share β (for bonus) and the consultant a share $1-\beta$ of output, with $\beta \in [0,1]$. If instead the

[8] We restrict α to lie in $[0,1]$ to prevent budget-breaking schemes in which a player acquires a negative share of output. We impose the same restriction on the share β defined below. Such restriction prevents incentives to sabotage output and is standard in the literature (see, for example, Eswaran and Kotwal, 1984).

producer receives help from the consultant, output is split between the consultant and the producer according to shares α and $1-\alpha$. The key distinction vis-a-vis ex-post contracts is that ex-ante contracts may condition payments on an additional event: whether or not the producer solves the problem on her own. Since there are only three possible outcomes, an ex-ante contract of the form $\langle \omega, \alpha, \beta \rangle$ is without loss.

Note that ex-post contracting is a special case of ex-ante contracting. Indeed, any ex-post contract $\langle w, \alpha \rangle$ can be replicated by an ex-ante contract that sets $\langle \omega, \alpha, \beta \rangle = \langle w, \alpha, 1-w \rangle$. Equivalently, ex-post contracting is formally identical to ex-ante contracting with the additional restriction that $\beta = 1-\omega$. It is worth pointing out that since ex-ante contracts are more general, teams are always weakly better off, by assumption, when they use such contracts. Therefore, our model is not suitable for providing a complete theory of when ex-ante, rather than ex-post, contracts are used in practice. What our model will deliver is an important advantage of ex-ante contracts as well as a characterization of their optimal shape.[9]

Although the main focus of our paper is the case in which the agents' knowledge z is unobservable, it is convenient to first analyze the full-information case. All omitted proofs are in Appendix A.

3 Benchmark: the first-best allocation

In this section we derive the first-best allocation and provide brief examples of full-information competitive equilibria that support it, including two examples that are of interest as they are robust to one-sided asymmetric information.[10]

3.1 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

[9] In practice, ex-ante contracting may entail costs relative to spot contracting, including the need for consultants to contract up front with a number of suitable producers and to cope with a potentially-random number of questions asked by those producers. Throughout the paper we abstract from such costs.

[10] The results in this section generalize the results of Garicano and Rossi-Hansberg (2004), who focus on a uniform type distribution F, and extend the results in Garicano and Rossi-Hansberg (2006) to the case in which skills are exogenous.

before producers have attempted their problems, with the understanding that consultants will attempt the unsolved problems of the producers they are matched with. Let $\mathcal{Z}, \mathcal{I} \subset [0,1]$ denote the sets of matched producers and unmatched (i.e. "self-employed") producers, respectively, and let $\mathcal{M} \subset [0,1]$ 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 some basic properties of any optimal allocation:

Lemma 1 *A surplus-maximizing allocation has the following features (except at most over a zero-measure subset of the type space):*

1. ***Positive-assortative matching.*** *Matching between producers and consultants is (strictly) 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 a consultant of type $m' \leq m$.*

2. ***Producer/consultant stratification.*** *The set $\mathcal{Z} \cup \mathcal{I}$ of producers lies below the set $\mathcal{M}$ of consultants.*

3. ***Within-producer stratification.*** *The set $\mathcal{Z}$ of matched producers lies below the set $\mathcal{I}$ of self-employed producers.*

Intuition for each part is as follows. Positive-assortative matching is optimal since the output function $\Pi(m, z) = n(z)m$ is supermodular and so a higher consultant knowledge is best exploited when applied to a harder problem. Producer/consultant stratification is optimal because a higher knowledge is more valuable in the hands of a consultant (who applies her knowledge to $\frac{1}{h}$ problems) than in the hands of a producer (who applies her knowledge to one problem only). Within-producer stratification is optimal because, provided the planner wishes consultants to attempt only a fraction of the producers' unsolved problems, it is best that consultants attempt the problems that are, on average, easiest to solve. These are precisely the problems that come from producers with the lowest types.

From this Lemma we immediately obtain:

Corollary 1 *A 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:*

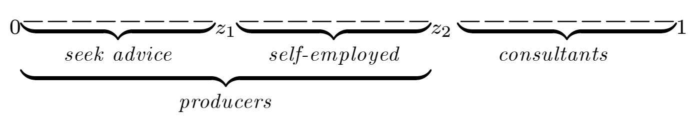
A. *Agents in $[0, z_1]$ are matched producers who seek advice if needed.*

B. *Agents in* (z_1, z_2) *are self-employed producers who never seek advice.*

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

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

The diagram below summarizes parts A-C.



In order to match producers and consultants, the planner selects a function $M : \mathcal{Z} \longrightarrow \mathcal{M}$ that assigns to each matched producer z a consultant $m = M(z)$. Positive-assortative matching means that M is an increasing function.[11]

When selecting M, the planner must satisfy a simple resource constraint. Fix the producer cutoff type z_1. For any type $z \in \mathcal{Z}$, the total time required to consult on the problems left unsolved by producers in the interval $[z, z_1]$ must equal the total time available to consultants in the interval $[M(z), 1]$:

$$h \int_z^{z_1} (1-t) dF(t) = \int_{M(z)}^1 dF(m) \text{ for all } z \leq z_1. \quad (1)$$

For any given z_1, there is a unique increasing function $M(z)$ that satisfies the above constraint, namely, the function satisfying $M(z_1) = 1$ and

$$M'(z) = h(1-z)\frac{f(z)}{f(M(z))} \text{ for all } z \leq z_1, \quad (2)$$

which follows from differentiating (1) with respect to z. Denote such function $M(z; z_1)$. Note that the consultant cutoff type z_2 is also uniquely determined by z_1, namely, $z_2 = M(0; z_1)$.

[11] Note that the literal interpretation of such pointwise matching is that any given positive-measure interval $(z-\varepsilon, z+\varepsilon)$ in $\mathcal{Z}$ is matched with the positive-measure interval $(M(z-\varepsilon), M(z+\varepsilon))$ in $\mathcal{M}$, with the function M satisfying the time budget constraint

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

where the L.H.S. is h times the total mass of problems left unsolved by producers in $(z-\varepsilon, z+\varepsilon)$ and the R.H.S. is the total mass of consultants in $(M(z-\varepsilon), M(z+\varepsilon))$.

Equipped with Corollary 1 and the matching function $M(z; z_1)$, we can state the planner's problem as a function of z_1 alone:

$$\max_{z_1} \int_0^{z_1} M(z; z_1) dF(z) + \int_{z_1}^{M(0;z_1)} z dF(z) \tag{I}$$

$$s.t.$$

$$z_1 \leq M(0; z_1). \tag{3}$$

The objective is equal to social surplus. The first term is the total mass of solved problems among those problems initially attempted by the matched producers in $\mathcal{Z} = [0, z_1]$ (indeed, the probability that a problem initially attempted by $z \in \mathcal{Z}$ is ultimately solved is $z + (1-z) \frac{M(z;z_1)-z}{1-z} = M(z; z_1)$). The second term in the objective is the total mass of problems solved by the self-employed producers in $\mathcal{I} = [z_1, M(0; z_1)]$. Finally, constraint 3 indicates that, per Corollary 1, z_1 cannot exceed $z_2 = M(0; z_1)$.

Let z_1^* denote a solution to the planner's problem (a function of both F and h). For notational simplicity, let $M^*(z) \equiv M(z; z_1^*)$ denote the corresponding optimal matching function and let $Z^*(m)$ denote the inverse of $M^*(z)$ (namely, consultant m is optimally matched with producer $z = Z^*(m)$).

Lemma 2 characterizes the solution to the planner's problem:

Lemma 2 *The planner's problem (I) has a unique solution z_1^*, which is strictly positive. Letting $z_2^* = M(0; z_1^*)$ denote the associated optimal value of z_2, this unique solution takes one of two forms, depending on whether or not having self-employed producers is efficient:*

1. *$z_1^* < z_2^*$ (i.e. $\mathcal{I} \neq \varnothing$) and*
$$\frac{1}{h} - z_2^* = \int_{z_2^*}^1 n(Z^*(t))\, dt. \tag{4}$$

2. *$z_1^* = z_2^*$ (i.e. $\mathcal{I} = \varnothing$) and*
$$\frac{1}{h} - z_2^* \geq \int_{z_2^*}^1 n(Z^*(t))\, dt. \tag{5}$$

Moreover, for any given F: (a) there exists a unique number $h_0(F) \in (0, 1)$ such that $z_1^* < z_2^*$ if and only if $h > h_0(F)$; and (b) z_1^* is decreasing in h.

Intuition is as follows. Part 1 corresponds to an interior solution for z_1^* and equation (4) describes the corresponding first-order condition. This first-order condition has a simple interpretation. Consider a marginal increase in z_1^* that has the effect, owing to the planner's resource constraint, of transforming one more agent (namely, agent z_2^*) into a consultant. The L.H.S. measures the net gain along the extensive margin: on the one hand, the highest-type consultant is now free to solve $\frac{1}{h}$ unsolved problems of types marginally above z_1^* (who were original self-employed); on the other hand, agent z_2^* no longer produces her original level of surplus z_2^*. The R.H.S. measures the combined loss along the intensive margin: every single producer z below z_1^* is now matched with a marginally worse consultant (since every consultant is matched with a marginally better producer), which in turn lowers the output of producer z's team by $\frac{\partial}{\partial m}\Pi(M^*(z), z) = n(z)$. Part 2 corresponds to a corner solution for z_1^* and inequality (5) states that, when increasing z_1 the net gain along the extensive margin is no lower than the combined loss along the intensive margin.

Finally, the last part of the Lemma implies that the efficient set of matched producers grows as the helping cost h falls. The reason is that reducing h increases the productivity of consultants (who can now consult on more problems) relative to producers (who, regardless of h, work on a single problem each).

3.2 Competitive equilibrium and one-sided adverse selection

When there is no asymmetric information, the first welfare theorem implies that the first best can be attained in a decentralized competitive equilibrium. In this section, we consider simple examples of such decentralizations. The goal of the section is three-fold. First, we characterize the agents' equilibrium payoffs in the absence of information asymmetries. Second, we show that, absent information asymmetries, multiple institutional arrangements support the first best. Third, we show that when information asymmetries are at most one-sided the first best can be implemented with simple institutional arrangements.

We begin by describing a simple (full-information) competitive arrangement. Assume free entry of risk-neutral competitive firms. Firms have no fixed costs and, upon entering the market, hire one consultant of a given type m and $n(z) = \frac{1}{h(1-z)}$ producers of a given type z.[12] Firms are the residual claimants of all output. Since agents are risk neutral and their types are known, we may assume without loss that firms pay their consultants and producers non-contingent, type-specific wages. Let $w_1(m)$ and $w_2(z)$ denote the wages received by a

[12] The literal interpretation is that the firm hires a small interval of producers $(z-\varepsilon_1, z+\varepsilon_1)$ and a small interval of consultants $(m-\varepsilon_2, m+\varepsilon_2)$ such that the mass of the interval of producers is $n(z)$ times the mass of the interval of consultants.

type m consultant and a type z producer, respectively. The profits of a firm are

$$\Pi(m,z) - w_1(m) - n(z)w_2(z)$$

(recall that $\Pi(m,z) = n(z)m$ is the firm's ex-ante gross profit).

Definition 1 (Competitive Equilibrium) *A competitive equilibrium consists of wage schedules $w_1^*, w_2^* : [0,1] \longrightarrow \mathbb{R}$, sets $\mathcal{Z}$, $\mathcal{I}$, $\mathcal{M}$ (which partition the set of types $[0,1]$ into matched producers, self-employed producers, and consultants, respectively), and a feasible matching function $M : \mathcal{Z} \longrightarrow \mathcal{M}$ such that:*

1. Firms maximize profits (while earning zero profits).

2. A firm that hires type z producers optimally hires a type $m = M(z)$ consultant.

3. Agents earn wages no lower than their outside options.

4. The markets for producers and consultants clear.

We now characterize the equilibrium wage schedules, which correspond to the equilibrium payoff schedules of matched producers and consultants. (Self-employed producers simply earn their outside options z). Note that, in equilibrium $\mathcal{Z} = [0, z_1^*]$, $\mathcal{I} = (z_1^*, z_2^*)$, and $\mathcal{M} = [z_2^*, 1]$, which correspond to the efficient occupational choices, and matching is given by the first-best function $M^*(z) \equiv M(z; z_1^*)$ (with inverse $Z^*(m)$).

For brevity, we focus on our main case of interest in which the first best calls for a positive mass of self-employed producers ($\mathcal{I} \neq \varnothing$). In Appendix B we consider the case in which $\mathcal{I} = \varnothing$.[13]

Remark 1 *Suppose the first best prescribes self-employed producers. In a full-information competitive equilibrium:*

A. $w_1^*(m) = z_2^* + \int_{z_2^*}^{m} \underbrace{n\left(Z^*(t)\right)}_{\frac{d}{dt}w_1^*(t)} dt$ *for all* $m \in \mathcal{M}$.

B. $w_2^*(z) = M^*(z) - \frac{1}{n(z)} w_1^*(M^*(z))$ *for all* $z \in \mathcal{Z}$.

Part A tells us that a consultant's equilibrium payoff $w_1^*(m)$ grows with her knowledge m at a rate equal to the full marginal contribution of her knowledge to firm profits, which in

[13] As shown in the appendix, the two cases differ only in how the equilibrium payoff of the marginal consultant z_2^* is determined.

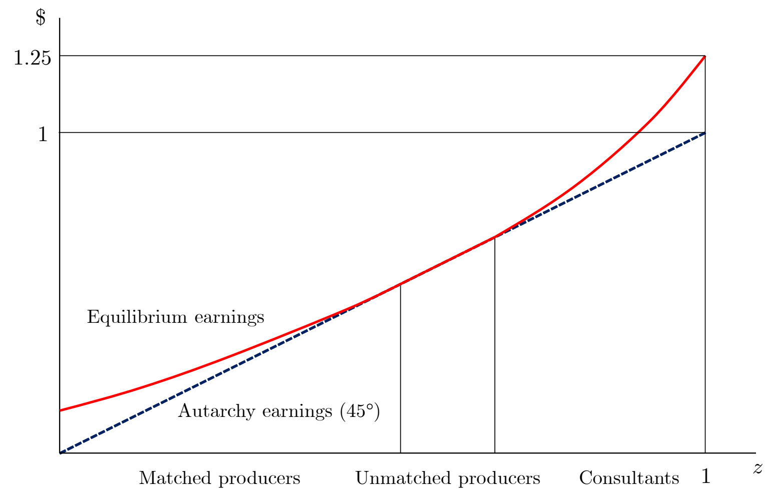


Figure 1: Full-information payoffs (for F uniform and $h = 0.8$)

turn equals the number of producers $n(Z^*(m))$ the consultant is matched with. This result follows from the fact that consultant knowledge is public information and there is perfect competition for such knowledge. Moreover, part A tells us that the lowest-type consultant must earn her autarky payoff z_2^*, which is simply a consequence of equilibrium payoffs being continuous in type. Part B tells us that, since firms earn zero profits in equilibrium, a producer's wage must equal the expected output per producer net of consultant wages.

Figure 1 illustrates the equilibrium payoffs. The agents who obtain the highest gains from trade are those at the extremes of the type space. On the one hand, the agents with the highest knowledge levels benefit from being able to leverage their knowledge by serving as consultants. On the other hand, the agents with the lowest knowledge levels benefit from being able to pose problems that are relatively easy to solve. In contrast, agents with intermediate knowledge levels provide limited value as both producers (since they pose problems that are relatively hard to solve) and as consultants (since their knowledge is low relative to that of other consultants). For this reason, agents with intermediate knowledge obtain particularly low (even zero) gains from trade.

Consultant and referral markets

It is worth noting that when information is symmetric there are a variety of ways to attain the first best by means of a competitive market. Such decentralizations vary in how the residual claim of output is allocated, but they all deliver identical expected payoffs.[14] Two simple examples of economic interest follow:

Example 1 ***A consultant market.*** *Consider a variation of the original competitive arrangement in which each producer owns a firm that handles the single problem of that producer. The producer/owner has the option of purchasing, upon failing to solve her problem, h units of consultant time at a type-specific fixed wage* $w_1(m)$ *(per unit of time). In equilibrium, the price per unit of consultant time for a type* m *consultant must be precisely equal to* $w_1^*(m)$*, as such price ensures that a producer of type* $Z^*(m) \in \mathcal{Z}$ *optimally hires a consultant of type* $m \in \mathcal{M}$*. Such decentralization is of interest because it is robust to producer types being private information (namely, the only information required to support the first best is the type of each consultant* $m \in \mathcal{M}$*.)*

Example 2 ***A referral market.*** *Consider a variation of the original competitive arrangement in which each consultant owns a firm that consults on* $\frac{1}{h}$ *unsolved problems of type* z *producers and pays each of those producers a type-specific price* $p(z)$*. (Equivalently, the consultant's firm may hire* $n(z)$ *producers up front at a type-specific wage* $w_2(z)$*). The ex-ante expected payoff earned by a matched producer* $z \in \mathcal{Z}$ *is then* $z + (1-z)p(z)$*. In equilibrium, such payoff must be precisely equal to* $w_2^*(z)$ *(and so* $p(z) = \frac{w_2^*(z)-z}{1-z}$*) which ensures that a consultant of type* $M^*(z) \in \mathcal{M}$ *optimally consults on the problems of producers of type* $z \in \mathcal{Z}$*. Such arrangement is of interest because it is robust to consultant types being private information (namely, the only information required to support the first best is the type of each producer* $z \in \mathcal{Z}$*.)*

The above are examples of simple institutional arrangements that achieve efficiency provided the informational asymmetries are at most one-sided. In general, in bilateral relationships, letting the party with private information be the residual claimant supports efficiency. The examples illustrate that such logic extends to the present two-sided market. As long as the market is set up so that prices are based on the observable type, equilibrium prices will induce the side of the market with private information to select the efficient match. The examples also highlight that, even holding the underlying trade constant, markets may organize in different ways if information asymmetries differ across them.

[14] Indeed, in every such decentralization the expected payoff of the marginal consultant z_2^* must equal her autarky payoff and consultant payoffs must grow with consultant knowledge at a rate equal to the full marginal value of that knowledge. Producers in turn obtain the remaining expected profits net of consultant payoffs.

Notably, as we show in section 4 below, when there is two-sided asymmetric information, granting consultants full residual claim of output will be *required* for efficiency whenever consultant time is scarce (i.e., whenever the first best calls for self-employed producers), leading to a unique family of ex-ante contracts that implements the first best.

4 Two-sided adverse selection: optimal contracts

We now turn to the case in which the agents' knowledge levels are private information. In order to obtain more favorable contractual terms, agents may wish to misrepresent their knowledge. Consultants, on the one hand, may want to overstate their knowledge in order to exaggerate their contribution to team output. Producers, on the other hand, may want to either understate or overstate their knowledge depending on whether they contract ex-post (in which case higher knowledge is undesirable) or ex-ante (in which case higher knowledge is desirable).

We are interested throughout in characterizing the family, or families, of contracts

$$\langle \omega(z), \alpha(z), \beta(z) \rangle_{z \in [0, z_1^*]}$$

that implement the first best. In what follows, to avoid notational clutter, let $M(\cdot)$ denote the first-best matching function $M^*(\cdot)$ and let $Z(\cdot)$ denote its inverse $Z^*(\cdot)$. In addition, let $\mathcal{Z} = [0, z_1^*]$, $\mathcal{I} = (z_1^*, z_2^*)$, and $\mathcal{M} = [z_2^*, 1]$.

Formally, we are interested in truth-telling mechanisms of the form $\langle \omega(z), \alpha(z), \beta(z), M(z) \rangle_{z \in \mathcal{Z}}$. A mechanism works as follows. After agents sort into occupations, they are invited to report their types to an abstract mechanism designer, with producers invited to report a type in $\mathcal{Z} \cup \mathcal{I}$ and consultants invited to report a type in $\mathcal{M}$. First, if a producer reports a type $z' \in \mathcal{Z}$ she is matched with consultant $M(z')$ under the bilateral contract $\langle \omega(z'), \alpha(z'), \beta(z') \rangle$. Second, if a producer reports a type $z' \in \mathcal{I}$ she is left unmatched (i.e. becomes self employed). Third, if a consultant reports type $m' \in \mathcal{M}$ she is matched with $n(z')$ producers of type $z' = Z(m')$ under bilateral contracts $\langle \omega(z'), \alpha(z'), \beta(z') \rangle$. Finally, any agent who does not participate in the mechanism is left unmatched.

We say that a family of contracts $\langle \omega(z), \alpha(z), \beta(z) \rangle_{z \in \mathcal{Z}}$ implements the first best if the mechanism $\langle \omega(z), \alpha(z), \beta(z), M(z) \rangle_{z \in \mathcal{Z}}$ induces truth telling once agents have sorted into their efficient occupations and, in addition, the mechanism induces agents to select their efficient occupations up-front.

The expected payoffs for players who participate in the mechanism are as follows:[15]

1. A producer of type $z \in [0,1]$ who reports type $z' \in \mathcal{Z}$ receives

$$R(z,z') = \omega(z') + z\beta(z') + (M(z') - z)(1 - \alpha(z')),$$

   where z is the probability that the producer solves her problem on her own, and $M(z') - z = (1-z)\frac{M(z')-z}{1-z}$ is the probability that she seeks advice from consultant $M(z')$ and manages to solve her problem upon seeking such advice.

2. A producer of type $z \in [0,1]$ who reports type $z' \in \mathcal{I}$ receives her self-employment payoff z.

3. A consultant of type $m \in [0,1]$ who reports type $m' \in \mathcal{M}$ receives

$$S(m,m') = n(z')\left[-\omega(z') + z'(1-\beta(z')) + (m - z')\alpha(z')\right] \text{ for } z' = Z(m'),$$

   where z' is the probability that each producer z' she is matched with solves her problem on her own, and $m - z' = (1-z')\frac{m-z'}{1-z'}$ is the probability that producer z' requires advice from consultant m and solves her problem conditional on receiving such advice.

For future reference, note that both $R(z,z')$ and $S(m,m')$ are linear in true type, with

$$R_z(z,z') = \alpha(z') + \beta(z') - 1 \text{ and } S_m(m,m') = \alpha(Z(m'))n(Z(m')). \tag{6}$$

Accordingly, R has increasing differences in z and $(\alpha(z') + \beta(z') - 1)$ and S has increasing differences in m and $\alpha(Z(m'))n(Z(m'))$.

With slight abuse of notation, let $R(z) = R(z,z)$ and $S(m) = S(m,m)$ denote, respectively, the payoffs of a producer in $\mathcal{Z}$ and a consultant in $\mathcal{M}$ who report their types truthfully. Note, finally, that R and S satisfy the condition that team surplus $(\Pi(M(z),z) = n(z)M(z))$ is fully allocated within each team:

$$S(M(z)) + n(z)R(z) = n(z)M(z) \text{ for all } z \in \mathcal{Z}. \tag{TS}$$

[15] A non-participating producer of type z receives her self-employemnt payoff z and a non-participating consultant receives zero.

In what follows, we investigate four scenarios, according to whether or not the first best prescribes self-employed producers and according to whether or not contracts are restricted to be ex-post. We proceed by first considering both forms of contracting when the first best prescribes self-employed producers (Section 4.1) and then considering both forms of contracting when the first best does not prescribe self-employed producers (Section 4.2):

- When self-employed producers are called for (Section 4.1), spot contracts never achieve the first best (Proposition 1). In contrast, ex-ante contracts always achieve the first best via a unique arrangement with firm-like properties (Theorem 1).
- When self-employed producers are not called for (Section 4.2), spot contracts may or may not achieve the first best, depending on the distribution of types F and the value of h. When F is uniform, for instance, spot markets are efficient (Proposition 2). Ex-ante contracts, as before, achieve the first best regardless of the underlying parameters (Theorem 2). In this case, there may be multiple arrangements that implement the first best.

4.1 Optimal contracting with self-employment

In this section we assume that the first best calls for a positive measure of self-employed agents (namely, h is sufficiently large), which is our main case of interest.

Incentive and occupational constraints

The following are necessary and sufficient conditions for a family of ex-ante contracts to implement the first best in the presence of self-employment:

First, the relevant incentive constraints *within* the efficient set of producers are met:

$$R(z) \geq R(z, z') \text{ for all } z, z' \in \mathcal{Z}. \tag{IC_1}$$

Second, the relevant incentive constraints *within* the efficient set of consultants are met:

$$S(m) \geq S(m, m') \text{ for all } m, m' \in \mathcal{M}. \tag{IC_2}$$

Third, and finally, every agent finds it optimal to select her first-best occupation:

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

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

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

We refer to these three inequalities collectively as the *occupational-sorting constraints*.[16]

We now simplify the above five constraints following standard steps. We begin with the incentive constraints (IC_1) and (IC_2). Making use of the Envelope Theorem (e.g. Milgrom and Segal, 2002, Theorem 2), together with the derivatives of R and S in (6), these constraints are each reduced to two simple conditions:

On the one hand, the producer incentive constraints (IC_1) are met if and only if

$$R(z) = R(0) + \int_0^z \underbrace{[\alpha(t) + \beta(t) - 1]}_{R'(t)} dt \text{ for all } z \in \mathcal{Z}, \quad (E_R)$$

$$R'(z) = \alpha(z) + \beta(z) - 1 \text{ is nondecreasing in } z, \quad (M_R)$$

where (E_R) is the usual envelope condition that captures local incentive constraints and (M_R) is the relevant monotonicity condition that ensures that all incentive constraints are met globally (across producers in $\mathcal{Z}$).

On the other hand, the consultant incentive constraints (IC_2) are met if and only if

$$S(m) = S(z_2^*) + \int_{z_2^*}^m \underbrace{\alpha(Z(t))n(Z(t))}_{S'(t)} dt \text{ for all } m \in \mathcal{M}, \quad (E_S)$$

$$S'(m) = \alpha(Z(m))n(Z(m)) \text{ is nondecreasing in } m, \quad (M_S)$$

where, as before, (E_S) is the usual envelope condition that captures local incentive constraints and (M_S) is the relevant monotonicity condition that ensures that all incentive constraints

[16] The first inequality indicates that no matched producer wishes to become either a self-employed producer (and receive her autarky payoff z) or a consultant (and receive payoff $\max_{m'} S(z, m')$). The second inequality indicates that no consultant wishes to become either a self-employed producer (and receive her autarky payoff m) or a matched producer (and receive payoff $\max_{z'} R(m, z)$). The third inequality indicates that no self-employed producer wishes to become either a matched producer (and receive payoff $\max_{z'} R(z, z')$) or a consultant (and receive payoff $\max_{m'} S(z, m')$).

are met globally (across consultants in $\mathcal{M}$).

We now turn to the occupational-sorting constraints (i)-(iii). The following Remark, which relies on standard results, indicates that these constraints can be reduced to two equalities (those in (PC)) and two inequalities (those in (DD_R) and (DD_S)).

Remark 2 *Suppose the first best prescribes self-employment ($\mathcal{I} \neq \emptyset$) and suppose the family of contracts $\langle \omega(z), \alpha(z), \beta(z) \rangle_{z \in \mathcal{Z}}$ satisfies (IC_1) and (IC_2). Then, that family of contracts satisfies the occupational-sorting constraints (i)-(iii) if and only if:*

1. *The agents' equilibrium payoffs are continuous at z_1^* and z_2^*:*

$$R(z_1^*) = z_1^*, \;\; S(z_2^*) = z_2^*. \tag{PC}$$

2. *The directional derivatives of R and S satisfy:*

$$R'(z_1^*-) \leq 1, \tag{DD_R}$$

$$1 \leq S'(z_2^*+), \tag{DD_S}$$

(where $R'(z_1^*-)$ denotes the left-hand derivative of R at z_1^* and $S'(z_2^*+)$ denotes the right-hand derivative of S at z_2^*).

Intuition is as follows. The agents' equilibrium payoffs are given by the function

$$V(z) = \begin{cases} R(z) & \text{for all } z \in \mathcal{Z}, \\ z & \text{for all } z \in \mathcal{I}, \\ S(z) & \text{for all } z \in \mathcal{M}. \end{cases}$$

Note that V is the upper envelope of the payoffs that agents can obtain under the family of available contracts, and, under any such contract, payoffs are continuous in type. It follows that V must be continuous in type, which delivers (PC). Moreover, it follows that at every point z the directional derivatives of V must satisfy $V'(z-) \leq V'(z+)$, which delivers (DD_R) and (DD_S).[17]

[17] See Milgrom and Segal (2002), Theorem 1. Indeed, if instead the payoff $R(z)$ of matched producers was steeper than the payoff z of self-employed producers at the cutoff type z_1^*, then agents marginally to the left of z_1^* would prefer to be self-employed. Similarly, if the payoff z of self-employed agents was steeper than the payoff of consultants at the cutoff type z_2^*, then agents marginally to the right of z_2^* would prefer to be self employed.

Conversely, (DD_R) and (DD_S) together with the monotonicity conditions ensure that $R(z)$ is (weakly) flatter than z and that $S(z)$ is (weakly) steeper than z. Accordingly, it suffices that the occupational sorting constraints are met at the occupational boundaries z_1^* and z_2^*, which is guaranteed by (PC).

In summary, the relevant constraints are the within-occupation incentive constraints (captured by (E_R), (M_R), (E_s), and (M_s)) and the occupational-sorting constraints (captured by (PC), (DD_R), and (DD_S)).

A necessary condition for optimality: consultants as residual claimants

We now prove a result that will become a central building block for the results that follow:

Lemma 3 *Suppose the first best prescribes self-employment ($\mathcal{I} \neq \varnothing$). Then, the family of contracts $\langle \omega(z), \alpha(z), \beta(z) \rangle_{z \in \mathcal{Z}}$ implements the first best **only if** $\alpha(z) = 1$ for every $z \in \mathcal{Z}$.*

Proof. One the one hand, from Lemma 2, the planner's first-order condition implies that

$$\frac{1}{h} - z_2^* = \int_{z_2^*}^{1} n(Z(t))dt. \qquad (a)$$

On the other hand, the consultant envelope condition (E_S), combined with the fact that $S(z_2^*) = z_2^*$, implies that

$$S(1) - z_2^* = \int_{z_2^*}^{1} \alpha(Z(t))n(Z(t))dt. \qquad (b)$$

From (TS), and the fact that $R(z_1^*) = z_1^*$, it follows that $S(1) = \frac{1}{h}$. Consequently, given that $\alpha(z) \in [0, 1]$ for all z, equations (a) and (b) can only be simultaneously satisfied if $\alpha(z) = 1$ for almost every z. That $\alpha(z) = 1$ for every z then follows from the consultant monotonicity constraint (M_S). ■

Intuition is as follows. First, the highest consultant $m = 1$ must receive *all* gains from trade when contracting with her producers z_1^*, as such producers are held to their reservation payoffs z_1^*. Accordingly, the payoff of $m = 1$ is pinned down at the highly-attractive level $\frac{1}{h}$ (which is the highest possible payoff that any agent in the economy can hope to get). Granting $m = 1$ such a high payoff causes a problem for the mechanism designer: since consultant payoffs are necessarily continuous in type, many agents will in principle wish to become consultants and earn payoffs similar to that of $m = 1$. As a result, the designer's problem becomes, de facto, one of maximally deterring self-employed producers from entering the consulting occupation. In order to achieve maximal deterrence, the designer must ensure that as a consultant's type m falls, $S(m)$ falls at the highest possible rate consistent with incentive compatibility. This goal is achieved by forcing consultants to internalize the full marginal contribution of their knowledge, namely, setting $\alpha(Z(m)) \equiv 1$.

Note that granting consultants the full residual claim of the problems they consult on implies that $S'(m) \equiv n(Z(m))$ (which is equal to the full marginal contribution of consultant knowledge to team output). This observation, combined with the fact that $S\left(z_2^*\right)$ is pinned down at z_2^*, implies that the only way to achieve the first best in the presence of self-employment is to grant every consultant, and therefore every agent, an equilibrium payoff exactly equal to her full-information competitive payoff.

Ex-post (spot) contracting with self-employment

We begin with the case in which contracting takes place in a spot market, after producers have attempted their problems. Only producers who have failed participate in this market. Recall that, formally, ex-post (spot) contracting is a special case of ex-ante contracting in which $\beta(z)$ is restricted to equal $1 - \omega(z)$ for all $z \in \mathcal{Z}$.

The following result shows that spot contracting can never achieve the first best in the presence of self-employment:

Proposition 1 *Suppose the first best prescribes self-employment ($\mathcal{I} \neq \varnothing$). Then, the first best cannot be supported by spot contracting.*

Intuitively, the reason for this result is that, under spot contracting, producers who should be self-employed cannot be excluded from the market. To see why, recall that the marginal consultant z_2^* is matched with $n(0) = \frac{1}{h}$ producers of type $z = 0$, and those producers always fail to solve their $\frac{1}{h}$ problems. Note, moreover, that consultant z_2^* must become the full residual claimant of all such unsolved problems (per Lemma 3), yielding a total expected output $\frac{1}{h}z_2^*$, and yet must earn a low expected payoff equal to her autarky payoff z_2^* (per Remark 2). Accordingly, under ex-post contracting, such arrangement would require that the consultant pay a positive fixed price $w(0)$ to her producers. In that case, however, every self-employed producer with an unsolved problem would enter the market by pretending to have knowledge $z = 0$, namely, by selling the full residual claim of her problem, at a positive price, to consultant z_2^*.[18]

As we show below, ex-ante contracting prevents such deviation from self-employed producers by forcing them to commit up front to share their output when succeeding on their own. Since such output sharing is more costly for producers with higher types, it successfully keeps the most talented producers away from the market, as required in the efficient allocation.

Ex-ante contracting with self-employment

This section presents our main result: Theorem 1. This Theorem shows that, through an appropriate use of fixed wages and ex-ante output sharing, the first best can always be achieved. Moreover, the family of contracts that supports the first best is unique.

Theorem 1 *Suppose the first best prescribes self-employment ($\mathcal{I} \neq \varnothing$). A family of contracts implements the first best if and only if*

$$\alpha(z) = 1;\ \beta(z) = hS(M(z));\ \textit{and}\ \omega(z) = M(z) - hS(M(z))\ \textit{for all}\ z \in \mathcal{Z},$$

where $S(M(z)) = z_2^* + \int_{z_2^*}^{M(z)} n\left(Z\left(t\right)\right) dt$.

These contracts have the following structure (sketched in Figure 2):

Producers, on the one hand, have the option of either remaining self-employed or opting for a simple contract $\langle \omega(z), 1, \beta(z) \rangle$. Under such contract, a producer receives a guaranteed payment $\omega(z)$, which is decreasing in z, together with a bonus $\beta(z)$ upon succeeding on her own, which is increasing in z.[19] These contracts screen producers precisely because they become higher-powered as z grows. Indeed, as a producer's knowledge grows she is more willing to accept a lower wage in exchange for a higher contingent reward.[20]

[18] This result opens the possibility that, in spot markets, lemons in the middle of the type distribution are likely to benefit from asymmetric information (at the cost of agents in the extremes). Indeed, it is plausible that such a result would hold in a broad class of models where agents with unknown skill levels self-select into occupations. While deriving second-best equilibria is beyond the scope of the present paper, exploring such possibility is a promising avenue for future work. We are grateful to an anonymous referee for this observation.

[19] That $\omega(z)$ is decreasing follows from the fact that $\omega'(z) = M'(z)\left[1 - hn(z)\right] < 0$ for all $z > 0$. That $\beta(z)$ is increasing follows from the fact that $\frac{d}{dz}S(M(z)) = n(z)M'(z) > 0$.

[20] Note that in the limit as z converges upward to z_1^*, producers receive zero fixed wages and the full residual value of their self-produced output, which is precisely the form of compensation received by self-employed producers.

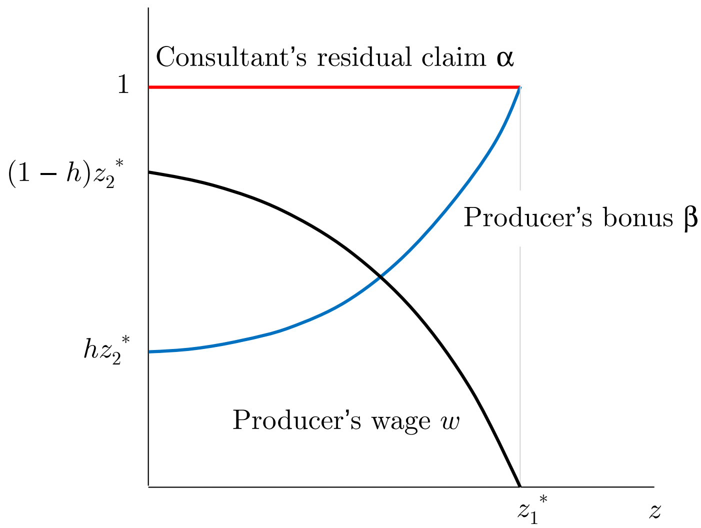


Figure 2: Optimal ex-ante contracts

Consultants, on the other hand, acquire the residual claim, net of bonus payments, of $n(z)$ problems, where z is the type of producers they opt to match with. In exchange, consultants offer each of the corresponding $n(z)$ producers an expected payment $\omega(z) + z\beta(z)$. These contracts screen consultants (who would prefer, other things equal, to be matched with more knowledgeable producers) by ensuring that the combined expected payment $n(z)\,[\omega(z) + z\beta(z)]$ to producers increases in z at a rate $n'(z)M(z)$ equal to the marginal contribution of producer knowledge to ex-ante team output, of which consultants are the residual claimants.

Note that consultants care only about the total expected payment $n(z)\,[\omega(z) + z\beta(z)]$ to producers, not about the way in which this payment is split between fixed and bonus payments. This property is what allows ex-ante contracts, with consultants as residual claimants, to simultaneously screen agents on both sides of the market.

The above contracts also guarantee that self-employed agents stay away from the market. That self-employed producers do not wish to become matched producers follows from the fact that being matched (under the pretense of having a type $z' \in \mathcal{Z}$) forces them to pay a share $1 - \beta(z')$ of output when succeeding on their own. Since such output sharing is, in expectation, more costly for producers with higher knowledge, it induces the producers with the highest knowledge levels (namely, self-employed producers) to remain unmatched. That self-employed producers do not wish to become consultants follows from the fact that, as consultants, they would become the full residual claimants of the unsolved problems they consult on, which they dislike because they are less capable, compared to existing consultants, of solving such problems.

Finally, that the above contracts are necessary to implement the first best follows from Lemma 3 (namely, $\alpha(z) \equiv 1$ is necessary to deter self-employed agents from becoming consultants) and the occupational-sorting constraints (namely, the marginal consultant z_2^* must earn her autarky payoff). These two requirements pin down the payoffs of all consultants, and, consequently, the payoffs of all matched producers. Moreover, screening matched producers requires that their equilibrium payoff $R(z)$ grows at rate $R'(z) = \beta(z)$, which uniquely pins down the bonus schedule.

Consultants as endogenous principals

Under the optimal contracts, each team can be interpreted as being formed by a principal (the consultant) and $n(z)$ employees (the producers). The principal pays each employee a fixed wage plus a performance bonus, and, net of these payments, she is the residual claimant of the team's output. Thus, the optimal contracts have a simple firm-like flavor that is reminiscent of the contracts proposed by Alchian and Demsetz (1972). In Alchian and Demsetz, specialized monitors endogenously become principals. In contrast, in the present setting, consultants who specialize in providing advice, and are not directly involved in production, endogenously become principals. Note that these principals are the most knowledgeable individuals in the economy.

4.2 Optimal contracting without self-employment

Here we consider the simpler case in which the surplus-maximizing allocation calls for no self-employment. This case arises, per Lemma 2, when offering advice is sufficiently inexpensive. As before, we are interested in characterizing the family, or families, of ex-ante contracts that implement the first best.

Incentive and occupational constraints (revisited)

The following are necessary and sufficient conditions for a family of contracts to implement the first best absent self-employment:

First, the incentive constraints (IC_1) and (IC_2) within the efficient sets of producers and consultants, respectively, must be met. As before, (IC_1) is equivalent to (E_R) and (M_R), and (IC_2) is equivalent to (E_S) and (M_S).

Second, every individual must find it optimal to select her first-best occupation:

$$R(z) \geq \max_{m' \in \mathcal{M}} S(z, m') \text{ for all } z \in \mathcal{Z}, \qquad (i')$$

$$S(m) \geq \max_{z' \in \mathcal{Z}} R(m, z') \text{ for all } m \in \mathcal{M}. \qquad (ii')$$

Third, and finally, the equilibrium payoff of each agent must be no smaller than her autarky payoff:

$$R(z) \geq z \text{ for all } z \in \mathcal{Z},\ S(m) \geq m \text{ for all } m \in \mathcal{M}. \qquad (IR)$$

(Note that this constraint was superfluous when the first best prescribed self-employment, as (IR) was already embedded in the corresponding occupational-sorting constraints.)

We now simplify the occupational-sorting constraints (i') and (ii'). The following Remark, which is a simpler version of Remark 2, indicates that these two constraints can be reduced to a single equality (that in (PC')) and a single inequality (that in (DD')):

Remark 3 *Suppose the first best prescribes no self-employment ($\mathcal{I} = \varnothing$) and suppose the family of contracts $\langle \omega(z), \alpha(z), \beta(z) \rangle_{z \in \mathcal{Z}}$ satisfies (IC_1) and (IC_2). Then, that family of contracts satisfies the occupational-sorting constraints (i') and (ii') if and only if:*

1. *The agents' equilibrium payoffs are continuous at z_1^*:*

$$R(z_1^*) = S(z_1^*). \qquad (PC')$$

2. *The directional derivatives of R and S satisfy:*

$$R'(z_1^*-) \leq S'(z_1^*+). \qquad (DD')$$

Intuition is akin to that of Remark 2. The agents' equilibrium payoffs are given by the function

$$V(z) = \begin{cases} R(z) & \text{for all } z \in \mathcal{Z}, \\ S(z) & \text{for all } z \in \mathcal{M}. \end{cases}$$

Incentive compatibility implies that V is continuos in type, which delivers (PC'), and, at every interior point z, $V'(z-) \leq V'(z+)$, which delivers (DD'). Conversely, from (DD') and the monotonicity conditions, it suffices that the occupational-sorting constraints are met at the single occupational boundary z_1^*, which is guaranteed by (PC').

In summary, the relevant constraints are now the within-occupation incentive constraints $((E_R)$, (M_R), (E_s), and $(M_s))$, the simplified occupational-sorting constraints $((PC')$ and $(DD'))$, and the participation constraint (IR).

Ex-post (spot) contracting without self-employment

When the first best does not call for self-employed producers, spot contracts may or may not support the first best, depending on the distribution of types F and the value of h. The reason why spot contracts may suffice for efficiency is that, absent self-employed producers, the designer has fewer constraints to satisfy as she is no longer restricted to grant the highest producer and lowest consultant their respective autarky payoffs.

Unfortunately, it is difficult to verify analytically whether or not the first best can be achieved for an arbitrary pair F, h. The reason is that the schedules w and α required for efficiency are uniquely pinned down by the local incentive constraints within the two occupations (namely, by the two envelope conditions (E_R) and (E_S)). Those schedules, in turn, have a complex shape, making it difficult to verify whether or not they meet all global incentive constraints, induce agents to sort into the desired occupations, and satisfy all participation constraints.[21]

For the remainder of this subsection, we focus on the special case in which F is uniformly distributed on $[0,1]$, for which we obtain a definite answer. Proposition 2 shows that for the uniform case, provided no self-employment is called for, the first best is implemented by a unique family of ex-post contracts.

Proposition 2 *Suppose the first best does not prescribe self-employment ($\mathcal{I} = \varnothing$). If knowledge is distributed uniformly on $[0,1]$, a family of spot contracts $\langle w(z), \alpha(z)\rangle_{z\in\mathcal{Z}}$ implements the first best if and only if, for all $z \in \mathcal{Z}$:*

1. $\alpha(z) = \frac{h(1-z)^2}{h(1-z)^2+1-M(z)}$, *which is increasing and convex in z and satisfies $\alpha(z_1^*) = 1$.*
2. $w(z) = h\left[\alpha(z)\pi(M(z), z) - \int_{z_1^*}^{M(z)} \alpha(Z(t))n(Z(t))\,dt - S(z_1^*)\right]$, *which is u-shaped and satisfies $w(z_1^*) \geq 0$.*[22]

[21] The general form for the necessary schedules α and w is presented in Appendix A (proof of Proposition 2).

[22] See Appendix A (proof of Proposition 2) for the value of the constant $S(z_1^*)$.

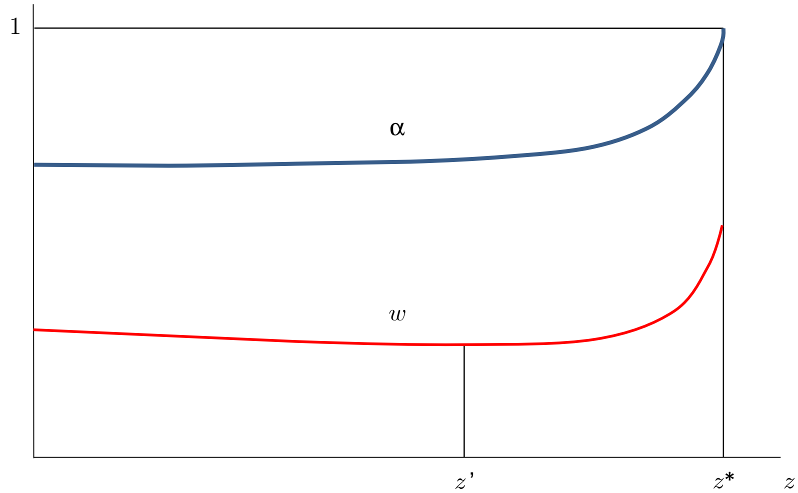


Figure 3: Spot contracts (for F uniform and $h = 0.3$); the u-shaped fee $w(z)$ reaches its minimum at $z' \simeq 0.53$

Moreover, M is concave with $M'(z) \equiv h(1-z)$.

Figure 3 depicts an example of α and w. Intuition is as follows. An increasing α helps sort agents on both sides. Higher consultant types prefer a higher residual claim since they are more likely to solve a given problem. Higher producer types prefer a lower residual claim since the problem they have is in expectation harder to solve. Furthermore, α can be shown to be convex. The reason is that consultants with high knowledge levels, who are in a unique position to consult on the difficult unsolved problems of the most knowledgeable producers, are awarded disproportionately large payoffs.[23]

Since α is increasing, one would in principle expect that more knowledgeable consultants pay a larger fixed fee in exchange for their larger share. This, however, is not the case. The u-shape of w follows from two countervailing effects: a matching effect and a rent-sharing effect. As a producer's type z grows she is matched with a more knowledgeable consultant (which, in and of itself, lowers the fixed fee she is willing to accept) and, simultaneously, she gives up a larger share of her rents to that consultant (which, in and of itself, raises the fixed fee she is willing to accept). For producers at the low end of the knowledge distribution the

[23] Indeed, the marginal value of a consultant's knowledge ($\pi_m = \frac{1}{h(1-z)}$) increases in the expected difficulty of the problem she faces at an increasing rate. Accordingly, in equilibrium, the incremental payoff received by a consultant who faces a marginally harder problem (i.e. $\frac{d}{dz}S(M(z))$) must also increase in the expected difficulty of that problem at an increasing rate. Furthermore, given that F is uniform, incentive compatibility dictates that $\frac{d}{dz}S(M(z))$ is precisely equal to her share $\alpha(z)$ (i.e. $\frac{d}{dz}S(M(z)) = \alpha(z)n(z)M'(z)$ and $M'(z) = \frac{1}{n(z)}$). It follows that $\alpha(z)$ must grow at an increasing rate.

matching effect is especially large (since the matching function is steep for small z) and the rent-sharing effect is especially small (since α is relatively flat), leading to a decreasing w. Precisely the opposite occurs for producers at the high end of the knowledge distribution, leading to an increasing w.

Note that, in an efficient spot market, consultants cannot be the full residual claimants of output. If that were the case, the fixed fee $w(z)$ would need to decline with z (as unsolved problems decline in value with z), which would in turn lead every producer to pretend to be of the lowest type $z = 0$.[24]

Currently, several spot markets are being set up using the internet to match 'seekers' of help with 'solvers.' An example is Innocentive (http://www.innocentive.com/), which as of August 2013 had more than 300,000 solvers. The appearance of such markets might be explained by the drop in communication costs (h) brought about by the internet. These exchanges, however, use only contingent pay (α) for those who succeed at solving a problem, and no fixed payments (w). Our results suggest that matching could be further improved with an appropriate mix of both types of payments.

Ex-ante contracting without self-employment

Theorem 2 extends our main result (Theorem 1) by showing that, absent self-employment, contracts with firm-like features remain efficient. Unlike the case with self-employment, the family of efficient contracts need not be unique. The reason is that, as noted above, the absence of self-employment grants the planner an additional degree of freedom: the equilibrium payoffs of the marginal producer/consultant.[25]

Theorem 2 *Suppose the first best does not prescribe self-employment ($\mathcal{I} = \varnothing$). The following family of contracts implements the first best:*

$$\alpha(z) = 1;\ \beta(z) = hS(M(z));\ \text{and } \omega(z) = M(z) - hS(M(z)) \text{ for all } z \in \mathcal{Z},$$

where $S(M(z)) = S(z_1^*) + \int_{z_1^*}^{M(z)} n\left(Z\left(t\right)\right) dt$.[26]

[24] As we show in Section 5.1, an intermediary monopolist can achieve efficiency while still granting zero residual claim, ex-post, to producers.

[25] For example, when F is uniform both the contracts in Proposition 2 and Theorem 2 are efficient.

[26] The constant $S\left(z_1^*\right)$ is given by

$$S\left(z_1^*\right) = \frac{1}{n\left(z_1^*\right) + 1}\left[n\left(z_1^*\right) - \int_{z_1^*}^{1} n\left(Z\left(t\right)\right) dt\right].$$

These contracts have the same qualitative properties as those in Theorem 1. Namely, consultants offer producers a fixed wage plus a success-contingent bonus, and, net of these payments, consultants are the residual claimants of all output. Moreover, as a producer's knowledge increases, her wage declines and her bonus rises.

The only differences vis-a-vis the contracts in Theorem 1 are: (1) the marginal producer/consultant receive a payoff $S(z_1^*)$ that lies (weakly) above their outside option z_1^*; and (2) the marginal producer z_1^* receives a wage (weakly) larger than zero and a bonus (weakly) smaller than 1 (in other words, the contract of the marginal producer no longer replicates her autarky arrangement).

5 Extensions

In this section we consider two extensions: (1) the case of a monopolistic intermediary who matches producers and consultants; and (2) the case of contracting absent contingent pay for consultants.

5.1 A monopolistic intermediary

In this section we consider a simple variation of the model in which all matching between producers and consultants is controlled by a monopolistic intermediary. For concreteness, we study an ex-post contracting environment in which, after agents have selected their occupations and producers have attempted their problems, the monopolist has the ability to purchase the residual claim of unsolved problems from producers and hire consultants to offer advice on those problems. We allow the monopolist to offer contracts consisting of both fixed and output-contingent payments.

Proposition 3 *Suppose a monopolist controls all matching between producers and consultants. Then, the monopolist's profits are uniquely maximized at the first-best allocation.*

This result illustrates how an intermediary who hires consultants and finds unsolved problems (in this case a stylized monopoly) can solve the failure of spot markets. Intuition is as follows. By purchasing unsolved problems for free and offering consultants output-contingent contracts that hold them to their outside options, the monopolist is able to keep all gains from trade for herself. As a result, she finds it in her interest to maximize such gains. Note that this outcome differs from the standard nonlinear-pricing monopoly outcome in which quality is distorted in order to reduce consumer information rents (e.g. Mussa and Rosen, 1978, Maskin and Riley, 1984). The reason is that, in the present setting with output-contingent contracts, the expected transfers to consultants depend on both the consultants' reported types, as usual, as well as on their *true* types (since a consultant with a higher type is paid a given output-contingent bonus with higher probability). By exploiting this fact, the monopolist is able to extract all information rents from its clients without distorting the allocation of resources.[27]

5.2 Market breakdown absent contingent pay

In this section we illustrate the key role played by success-contingent consultant pay (through either output-contingent contracts or straight transfers of problem ownership). Specifically, we assume that consultants can only be paid a fixed fee (independent of success), but we otherwise allow for general random allocations, including rationing on either side of the market. Note that this is the first time we consider random matching. The reason is that, up to this point, our goal was to support the first best, which (per Lemma 1) calls for deterministic matching.

Proposition 4 shows that, absent contingent pay, there is a full market breakdown regardless of the potential gains from trade:

Proposition 4 *Suppose consultants cannot receive success-contingent pay. Then, there is no competitive equilibrium with a positive level of trade.*

The intuition for this result is that, on the one hand, the expected earnings of a producer are increasing in her knowledge and, on the other hand, the expected earnings of a consultant are independent of her knowledge (owing precisely to the lack of contingent pay). As a result, the occupational-sorting incentives are fully misaligned: if an agent with a given knowledge level wishes to become a consultant, then all agents with lesser knowledge follow the same path.

Unlike in a standard lemons market (e.g. Akerlof, 1970), where unravelling from the top leads to various degrees of market failure (depending on the magnitude of the potential gains

[27] We are grateful to an anonymous referee for suggesting this extension.

from trade), in the present setting occupational deviations from the lowest types guarantee a full market breakdown.

To illustrate, consider the market for brain surgeons. That very good surgeons may not be differentially compensated from merely good surgeons appears to be a second-order consideration when supporting trade. The first-order challenge arises from the average Joe donning a white robe and offering to crack your head open. Accordingly, certification can be particularly valuable in environments in which residual ownership cannot be transferred and contingent contracts cannot be written. Fuchs and Garicano (2010) take the present result as their point of departure and show how a simple certification process partially restores efficiency.[28]

6 Conclusion

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%. Approximately half of this increase is the result of a growth in the share of professional and business services, such as computing, consulting, and legal services (Herrendorf, Rogerson, and Valentinyi, 2009). While production in the simplest manufacturing and agricultural sectors requires mostly labor and capital, production in the service sector is knowledge-intensive: capital equipment is largely irrelevant (as seen, for instance, in the law) and firms' success comes down strictly to their ability to deploy human capital appropriately.

In this paper, we study contractual arrangements in such knowledge-intensive industries. Leveraging expertise requires the use of knowledge-based hierarchies, where experts specialize in providing help to a number of non-experts. A crucial obstacle to such vertical specialization is asymmetric information – it is difficult to assess the knowledge levels of agents asking for, and those offering to, help. Can contracting resolve this problem?

We give a sharp answer to this question. In our model, there is a unique and simple arrangement that is guaranteed to organize knowledge efficiently. This arrangement is consistent with what we observe in practice: experts endogenously become residual claimants of output and pay a fixed wage and a performance bonus to the producers they are matched with. Moreover, producer incentives become higher-powered as their knowledge level grows. This arrangement simultaneously screens agents on both sides of the market because, on the

[28] See also early work by Leland (1979) and Shapiro (1986) on licensing.

one hand, consultants care only about their total payment to producers and, on the other hand, producers care about the way in which this total payment is split between fixed and bonus payments. Notably, these contracts are the only ones that deter self-employed agents from overstating their knowledge and becoming consultants, as they force them to internalize the full marginal value of their relatively-low knowledge.

This simple characterization of the optimal ex-ante contracts suggests a rationale for the organization of firms and the structure of compensation in the knowledge-intensive sector (e.g. legal, consulting, advertising, and private equity firms). A salient stylized fact of this sector is that agents with the highest expertise tend to be the residual claimants of income. In the manufacturing sector, in contrast, residual claimants tend to be the suppliers of capital, as opposed to the agents with the highest substantive knowledge about production.

Although spot market contracting is in general inefficient, as communication costs decrease such contracting might attain efficiency. This result may help explain the recent appearance of online spot exchanges for problem solving such as Innocentive.com.

We have focused throughout on contracting as a device to solve a problem of double-sided adverse-selection. In practice, firms and markets rely on various additional remedies including, notably, certification and reputation. A possible avenue for future work is studying how these different mechanisms interact. Finally, we have studied the value of ex-ante contracting while abstracting from the costs that such contracting is likely to entail relative to spot contracting, including the need for experts to contract up front with a number of suitable producers and to cope with a potentially-random number of questions asked by those producers. An additional possible avenue for future work is building a more comprehensive theory of the emergence of ex-ante contracting in the light of such costs.

7 Appendix A: omitted proofs

This section contains omitted proofs.

Proof of Lemma 1. Part 1, which follows from the (strict) supermodularity of Π, is proven in a more general framework by Eeckhout and Kircher (2012) (see their Proposition 1 and Section 3.3.4).

For part 2 we must show 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. We proceed by contradiction. Suppose the desired claim does not hold – namely, there is a positive-measure subset of consultants $\mathcal{M}_0$ that lies below a positive-measure subset of producers $\mathcal{Z}_0$. Select $\mathcal{M}_0$ and $\mathcal{Z}_0$ so that they have equal measures and they each fit in an interval of length $\varepsilon > 0$. Denote the measure of each of these sets Q. Note that no unsolved problem of a type z in $\mathcal{Z}_0$ is allocated to a consultant with a type m in $\mathcal{M}_0$ because that consultant's time would be wasted.

Now (1) transform all types in $\mathcal{Z}_0$ into consultants and allocate to them the unsolved problems initially allocated to the original consultants in $\mathcal{M}_0$ and (2) transform all types in $\mathcal{M}_0$ into producers and allocate their unsolved problems to the consultants that initially received the unsolved problems of the original producers in $\mathcal{Z}_0$ (provided such consultants exist and only up to the time constraint of those consultants). Note that change (1) increases social surplus by at least $\frac{1}{h}(\inf \mathcal{Z}_0 - \sup \mathcal{M}_0)Q$ (which is a lower bound on how many more problems are solved by the new consultants in $\mathcal{Z}_0$ relative to the old consultants in $\mathcal{M}_0$) and change (2) reduces social surplus by no more than $(\sup \mathcal{Z}_0 - \inf \mathcal{M}_0)Q$ (which is an upper bound on how many fewer problems are solved by the new producers in $\mathcal{M}_0$ relative to the old producers in $\mathcal{Z}_0$, which, given that some of those unsolved problems may be attempted by a consultant, constitutes in turn an upper bound on how many fewer problems attempted by the new producers in $\mathcal{M}_0$ are ultimately solved relative to the problems originally attempted by $\mathcal{Z}_0$). Since $\frac{1}{h} > 1$, the gain from (1) exceeds the loss from (2) for all small ε.

For part 3, let z_2 be such that almost all $z < z_2$ are producers. We must show that there is a cutoff $z_1 \leq z_2$ such that (almost) all producers $z \leq z_1$ seek advice and (almost) all $z \in (z_1, z_2)$ are self-employed producers. We proceed by contradiction. Suppose the desired claim does not hold – namely, there is a positive-measure subset of self-employed producers $\mathcal{Z}_0$ that lies below a positive-measure subset of matched producers $\mathcal{Z}_1$. Select $\mathcal{Z}_0$ and $\mathcal{Z}_1$ so that both subsets of producers leave the same mass of unsolved problems. Now swap the roles of $\mathcal{Z}_0$ and $\mathcal{Z}_1$. Since the unsolved problems of producers in $\mathcal{Z}_0$ are on average easier than those of producers in $\mathcal{Z}_1$, the associated consultants solve them with a higher probability while consuming the same amount of time, which increases social surplus. ■

Proof of Lemma 2. Let $\widehat{z}$ denote the largest value of z_1 such that constraint (3) is met and replace that constraint with the equivalent constraint $z_1 \leq \widehat{z}$.

The planner's first-order condition is

$$\int_0^{z_1^*} \frac{\partial}{\partial z_1} M(z; z_1^*) dF(z) + M(0; z_1^*) f\left(M(0; z_1^*)\right) \frac{\partial}{\partial z_1} M(0; z_1^*) + \left(1 - z_1^*\right) f(z_1^*) = \lambda,$$

where λ denotes the Lagrange multiplier for constraint $z_1 \leq \widehat{z}$, and the planner's complementary slackness condition is $\lambda \left[\widehat{z} - z_1\right] = 0$. Note, in addition, that

$$\frac{\partial}{\partial z_1} M(z; z_1) = -h\left(1 - z_1\right) \frac{f(z_1)}{f\left(M(z; z_1)\right)} \text{ for all } z_1, \tag{7}$$

which follows from differentiating constraint (1) with respect to z_1. The first-order condition therefore simplifies to[29]

$$-\int_{M(0;z_1^*)}^1 n(Z^*(m))dm - z_2^* + \frac{1}{h} = \lambda \frac{1}{h(1 - z_1^*) f(z_1^*)}.$$

Consequently, at the optimum, either: (1) $z_1^* < \widehat{z}$, $\lambda = 0$, and

$$z_2^* + \int_{z_2^*}^1 n(Z^*(m))dm = \frac{1}{h}; \tag{8}$$

or (2) $z_1^* = \widehat{z}$, $\lambda \geq 0$, and

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

As claimed in parts 1 and 2 of the Lemma.

That z_1^* is unique follows from equation (8) having a unique solution, which in turn

[29] To see why, note that from eq. (2) and integrating by substitution, $\int_{M(0;z_1^*)}^{M(x;z_1^*)} n(Z^*(m))dm = \int_0^x \frac{f(z)}{f\left(M(z;z_1^*)\right)} dz$ for all $x \leq z_1^*$, which in turn implies from eq. (7) that $\int_{M(0;z_1^*)}^{M(x;z_1^*)} n(Z^*(m))dm = -\frac{1}{h(1-z_1^*)f(z_1^*)} \int_0^x \frac{\partial}{\partial z_1} M(z; z_1^*) dF(z)$ for all $x \leq z_1^*$, and therefore $\int_{M(0;z_1^*)}^1 n(Z^*(m))dm = -\frac{1}{h(1-z_1^*)f(z_1^*)} \int_0^{z_1^*} \frac{\partial}{\partial z_1} M(z; z_1^*) dF(z)$.

follows from its L.H.S. being decreasing in z_1^*.[30] That z_1^* is a maximum, not a minimum, and is strictly positive, follows from noting that, in addition, the derivative of the planner's objective evaluated at $z_1 = 0$ is positive, namely,

$$M(0;0)f\left(M(0;0)\right)\frac{\partial}{\partial z_1}M(0;0) + f(0) =$$
$$(1-h)f(0) > 0,$$

where the equality follows from (7) and the fact that $M(0;0) = 1$.

Finally, we turn to the last part of the Lemma. Note that, making use of (7), and rearranging terms, the marginal contribution of z_1 to social surplus is

$$(1-z_1)\,f(z_1)\left\{1 - h\left[\int_{M(0;z_1)}^{1} n(Z(m;z_1))dm + M(0;z_1)\right]\right\} \qquad (9)$$

(where $Z(\cdot;z_1)$ is the inverse of $M(\cdot;z_1)$). Note that the expression in square brackets is greater than 1 for all $z_1 \in (0, \widehat{z}\,]$ (since $n(Z(m;z_1)) > 1$ for all $m \in [M(0;z_1), 1]$). It follows that, for any $z_1 > 0$, expression (9) is decreasing in h and can be either positive or negative depending on the value of $h \in (0,1)$. It follows that, whenever constraint (3) is slack, z_1^* is decreasing in h and therefore the desired cutoff $h_0(F) \in (0,1)$ exists. Moreover, when constraint (3) binds, z_1^* is decreasing in h owing to the time constraint (1) (i.e. when h falls a given set of consultants can help a larger set of producers). ■

Proof of Remark 1. Consider part A. Since gross output Π is differentiable in m, the function w_1^* must be a differentiable in m for all $m \in (z_2^*, 1)$ (and continuous for all $m \in [z_2^*, 1]$). We first show that $\frac{d}{dm}w_1^*(m) = n(Z(m;z_1^*))$ for all $m \in (z_2^*, 1)$. Consider the problem of a firm who has chosen to hire $n(z)$ producers of an arbitrary type $z \in (0, z_1^*)$ and

[30] To see why note that

$$\frac{d}{dz_1^*}\left[z_2^* + \int_{z_2^*}^{1} n(Z^*(m))dm\right] = \frac{d}{dz_1^*}M\left(0;z_1^*\right)\left(1 - n\left(Z^*(z_2^*)\right)\right) + \int_{z_2^*}^{1}\frac{d}{dz_1^*}n(Z^*(m))dm.$$

The first term on the R.H.S. is positive because $\frac{d}{dz_1^*}M\left(0;z_1^*\right) < 0$ and $n\left(Z^*(z_2^*)\right) = n(0) = \frac{1}{h} > 1$. The second term on the R.H.S. is also positive because $\frac{d}{dz_1^*}n(Z^*(m)) > 0$ for all m, as increasing z_1^* implies that each consultant is matched with producers of a higher type and so can be matched with a greater number of such producers.

must decide what type of consultant to hire:

$$\max_{m\in\left(z_2^*,1\right)} \quad \Pi(m,z) - w_1^*(m) - n(z)w_2^*(z).$$

The corresponding first-order condition (evaluated at $m = M^*(z)$) delivers the equilibrium slope of w_1:

$$\frac{d}{dm}w_1^*(M^*(z)) = \Pi_m(M^*(z),z) = n(z) \text{ for any } z \in \left(0, z_1^*\right).$$

Equivalently,

$$\frac{d}{dm}w_1^*(m) = n(Z^*(m)) \text{ for any } m \in \left(z_2^*, 1\right),$$

as desired.

Next, we show that $w_1^*(z_2^*) = z_2^*$. On the one hand, market clearing requires that $w_1^*(m) \geq m$ for all $m \in \mathcal{M}$ and $w_1^*(z) \leq z$ for all $z \in \mathcal{I}$ (otherwise agents would not sort into their efficient occupations). On the other hand, profit maximization requires that $w_1^*(z_2^*) \leq w_1^*(z_2^*-)$, where z_2^*- is a type lower than, and arbitrarily close to, z_2^* (otherwise no firm would hire a consultant of type z_2^* since hiring a consultant with a marginally lower knowledge instead would discontinuously raise profits). It follows that $z_2^* \leq w_1^*(z_2^*) \leq w_1^*(z_2^*-) \leq z_2^*-$ and therefore $w_1^*(z_2^*) = z_2^*$.

Consider part B. The zero-profit condition for a firm that hires producers of an arbitrary type $z \in \mathcal{Z}$ and a consultant of type $M^*(z)$ implies

$$n(z)M^*(z) - w_1^*(m) - n(z)w_2^*(z) = 0 \text{ for all } z \in \mathcal{Z}.$$

Upon rearranging terms, we obtain the desired value of $w_2^*(z)$. ■

Proof of Remark 2. Necessity ($\Rightarrow$). (PC) follows from the fact that the agents' equilibrium payoffs must be continuous in their type (a straightforward implication of local incentive compatibility). (DD_R) and (DD_S) follow from Theorem 1 in Milgrom and Segal (2002).

Sufficiency ($\Leftarrow$). Conditions (M_R), (M_S), (DD_R), and (DD_S) ensure that $R'(z) \leq 1$ and $S(z) \geq 1$. It therefore suffices that the occupational-sorting constraints are met at the occupational boundaries z_1^* and z_2^*, which in turn is guaranteed by (PC). ■

Proof of Proposition 1. Suppose toward a contradiction that the family of ex-post contracts $\langle w(z), \alpha(z)\rangle_{z\in\mathcal{Z}}$ implements the first best. First, note that consultant z_2^* is matched with $n(0) = \frac{1}{h}$ producers of type $z = 0$ and that team generates output $\Pi\left(z_2^*, 0\right) = \frac{1}{h}z_2^*$. It follows from (TS) that

$$S(z_2^*) + \frac{1}{h}R(0) = \frac{1}{h}z_2^*.$$

Second, recall that Remark 2 requires that $S(z_2^*) = z_2^*$ and Lemma 3 requires that $\alpha(0) = 1$. It follows that $R(0) = w(0)$ and so the above equality simplifies to

$$z_2^* + \frac{1}{h}w(0) = \frac{1}{h}z_2^*,$$

which implies that $w(0) > 0$.

Finally, given the existence of contract $\langle w(0), \alpha(0)\rangle = \langle w(0), 1\rangle$, every self-employed producer $z \in \mathcal{I}$, upon failing to solve her problem, would gain by pretending to be of type 0 and thus selling her unsolved problem to a consultant of type z_2^*, a contradiction. ■

Proof of Theorem 1. Necessity ($\Rightarrow$). Suppose the family of contracts $\langle \omega(z), \alpha(z), \beta(z)\rangle_{z\in\mathcal{Z}}$ implements the first best. First, from Lemma 3, we obtain $\alpha(z) = 1$ for all $z \in \mathcal{Z}$. Second, by differentiating (TS), namely, $R(z) = M(z) - \frac{1}{n(z)}S(M(z))$, with respect to z we obtain

$$R'(z) = M'(z)\,[1 - \alpha(z)] + hS(M(z)) \text{ for all } z \in \mathcal{Z}.$$

Since $R'(z) = \alpha(z) + \beta(z) - 1$ and $\alpha(z) = 1$, the above equality delivers $\beta(z) = hS(M(z))$ for all $z \in \mathcal{Z}$. Third, and finally, (TS) evaluated at the above values for $\alpha(z)$ and $\beta(z)$ implies that

$$\omega(z) + zhS(M(z)) = M(z) - h(1 - z)S(M(z)) \text{ for all } z \in \mathcal{Z},$$

which after rearranging terms delivers $\omega(z) = M(z) - hS(M(z))$ for all $z \in \mathcal{Z}$.

Sufficiency ($\Leftarrow$). We need to show that the family of contracts in the Theorem meets the seven relevant constraints, namely, (E_R), (M_R), (E_S), (M_S), (PC), (DD_R), and (DD_S). Note that $R(z) = M(z) - (1 - z)hS(M(z))$ (from (TS)) and so $R'(z) = hS(M(z))$.

(E_R) is met because $R'(z) = hS(M(z)) = \beta(z)$. (M_R) is met because $S(M(z))$ is increasing in z. (E_S) is met from the definition of $S(M(z))$. (M_S) is met because $S'(m) = n(Z(m))$ is increasing in m. (PC) is met because, on the one hand, from (TS),

$$R(z_1^*) = M(z_1^*) - h(1 - z_1^*)S(M(z_1^*)) = z_1^*,$$

where the second equality follows from the fact that $M(z_1^*) = 1$ and $S(1) = \frac{1}{h}$ (owing to the definition of S and Lemma 2, part 1), and, on the other hand, $S(z_2^*) = z_2^*$ by definition.

Finally, (DD_R) and (DD_S) are met because, on the one hand, $R'(z_1^*) = hS(1) = 1$, and, on the other hand, $S'(z_2^*) = n(Z(z_2^*)) > 1$. ■

Proof of Remark 3. Necessity ($\Rightarrow$). (PC') follows from the fact that the agents' equilibrium payoffs must be continuous in their type. (DD') follows from Theorem 1 in Milgrom and Segal (2002).

Sufficiency ($\Leftarrow$). Since payoffs exhibit increasing differences, (DD') and the monotonicity conditions (M_R) and (M_S) imply that it suffices that the occupational-sorting constraints are met at the single occupational boundary z_1^*, which in turn is guaranteed by (PC'). ■

Proof of Proposition 2. Necessity ($\Rightarrow$). Suppose the family of contracts $\langle w(z), \alpha(z)\rangle_{z\in\mathcal{Z}}$ implements the first best. First, by differentiating (TS) (namely, $S(M(z)) + n(z)R(z) = n(z)M(z)$) with respect to z we obtain

$$S'(M(z))M'(z) + n'(z)\left[R(z) - M(z)\right] = n(z)\left[M'(z) - R'(z)\right] \text{ for all } z \in \mathcal{Z}.$$

Using the definition of $R(z)$ and the facts that $R'(z) = \alpha(z) - w(z)$ and $S'(m) = \alpha(Z(m))n(Z(m))$ (which follow from the envelope conditions (E_R) and (E_S)), the above equality is equivalent, after rearranging terms, to

$$\alpha(z) = \frac{M'(z)(1-z)}{M'(z)(1-z) + 1 - M(z)} \text{ for all } z \in \mathcal{Z}.$$

Since F is uniform, $M'(z) = h(1-z)$ (from eq. (2)), which delivers part 1 of the Proposition.

Second, by combining (E_S) with the definition of $S(M(z))$ we obtain

$$w(z) = h\left[\alpha(z)\pi\left(M(z), z\right) - \int_{z_1^*}^{M(z)} \alpha(Z(t))n(Z(t))\,dt - S(z_1^*)\right] \text{ for all } z \in \mathcal{Z},$$

where the constant $S(z_1^*)$ uniquely satisfies (PC') (namely, $R(z_1^*) = S(z_1^*)$), as well as (TS) and (E_S) evaluated at z_1^*.[31]

Sufficiency ($\Leftarrow$). We need to show that the above schedules α and w meet the seven relevant constraints, namely, (E_R), (M_R), (E_S), (M_S), (PC'), (DD'), and (IR). The two

[31] (TS) and (E_S) imply, respectively, that $S(1) + n(z_1^*)S(z_1^*) = n(z_1^*)$ and $S(1) = S(z_1^*) + \int_{z_1^*}^1 \alpha(Z(t))n(Z(t))\,dt$, from which we obtain $S(z_1^*)$. Namely,

$$S(z_1^*) = \frac{1}{n(z_1^*)+1}\left[n(z_1^*) - \int_{z_1^*}^1 \alpha(Z(t))n(Z(t))\,dt\right].$$

envelope conditions (E_R) and (E_S) are met because these schedules imply that $R'(z) = \alpha(z) - w(z)$ and $S'(m) = \alpha(Z(m))n(Z(m))$, as required. The monotonicity condition (M_S) is met because $\alpha'(z) > 0$. Constraint (PC') is met by construction of the constant $S(z_1^*)$ (both the highest producer and lowest consultant earn this constant). Next, we numerically verify that the remaining three constraints (M_R), (DD'), and (IR) are met. We do this by noting that α and w, and therefore all payoffs, are continuous in h and by confirming that these three constraints are met for all h on a fine grid. We also derive the shapes of w and α numerically. (Program available upon request.) Finally, that $w(z_1^*) \geq 0$ follows from the (IR) constraint and the fact that $\alpha(z_1^*) = 1$, and that $M'(z) \equiv h(1-z)$ was noted above. ■

Proof of Theorem 2. We need to show that the family of contracts in the Theorem meets the seven relevant constraints, namely, (E_R), (M_R), (E_S), (M_S), (PC'), (DD'), and (IR). We begin by noting that the constant $S(z_1^*)$ in the Theorem uniquely satisfies

$$\underbrace{S(z_1^*) + \int_{z_1^*}^{1} n(Z(t))\,dt}_{S(1)} = n(z_1^*)\left[1 - S(z_1^*)\right]. \tag{10}$$

Since the L.H.S. is increasing in $S(z_1^*)$, the R.H.S. is decreasing in $S(z_1^*)$, and when $S(z_1^*) = z_1^*$ the L.H.S. is no larger than the R.H.S. (per the inequality in Lemma 2, part 2), it follows that $S(z_1^*) \geq z_1^*$.

We now show that each constraint is met. (E_R) is met because $R'(z) = hS(M(z)) = \beta(z)$. (M_R) is met because $S(M(z))$ is increasing in z. (E_S) is met from the definition of $S(M(z))$. (M_S) is met because $S'(m) = n(Z(m))$ is increasing in m. (PC') is met because (TS) evaluated at z_1^* (namely, $R(z_1^*) = M(z_1^*) - \frac{1}{n(z_1^*)}S(M(z_1^*))$) delivers

$$R(z_1^*) = 1 - \frac{1}{n(z_1^*)}S(1) = S(z_1^*),$$

where the first equality follows from the fact that $M(z_1^*) = 1$ and the second equality follows from the fact that $S(1) = n(z_1^*)[1 - S(z_1^*)]$ (from eq. (10)). (DD') is met because, on the one hand, $R'(z_1^*) = hS(1) \leq 1$ (where the inequality follows from the fact that, from eq. (10), $S(1) = n(z_1^*)[1 - S(z_1^*)]$ and $S(z_1^*) \geq z_1^*$), and, on the other hand, $S'(z_1^*) = n(Z(z_1^*))$ > 1. Finally, (IR) follows from the following facts: $R(z_1^*) = S(z_1^*) \geq z_1^*$; $R'(z_1^*) = hS(M(z))$ $\leq hS(1) \leq 1$ for all $z \in \mathcal{Z}$; and $S'(m) = n(Z(m)) > 1$ for all $m \in \mathcal{M}$. ■

Proof of Proposition 3. It suffices to show that the monopolist can implement the first best and, simultaneously, obtain all gains from trade (namely, hold all agents to their self-employment payoffs). Assume that all agents sort into their first-best occupations (we will show that, given the contracts that follow, such choices are indeed optimal). First, let the monopolist offer to purchase the residual claim of all unsolved problems from producers $z \in [0, z_1^*]$ for free while at the same time asking those producers to truthfully reveal their types. As producers have nothing to lose, they agree to both. Second, let the monopolist offer consultants a menu of contracts $\langle \tau_0(m), \tau_1(m), Z^*(m) \rangle_{m \in [z_2^*, 1]}$, where $\tau_0(m)$ is a fixed payment, $\tau_1(m)$ is a bonus per problem solved, and $Z^*(m)$ is the producer match enjoyed by a consultant who reports to be of type m. Now set

$$\tau_0(m) = Z^*(m) \text{ and } \tau_1(m) = h(1 - Z^*(m)) \text{ for all } m \in [z_2^*, 1] .$$

Note that every consultant $m \in [z_2^*, 1]$ earns her self-employment payoff m regardless of her report. As a result, all consultants agree to report their types truthfully. Since both producers and consultants report their types truthfully, the monopolist can indeed implement the first-best match $Z^*(\cdot)$. Finally, note that under such contracts, all agents earn their self-employment payoffs regardless of their career choices and reports, and therefore they find it optimal to select their first-best occupations from the outset. ■

Proof of Proposition 4. Assume toward a contradiction that there exists an equilibrium with a positive mass of consultants. Let $\mathcal{M}^*$ and $\mathcal{Z}^*$ denote, respectively, the equilibrium set of consultants and the equilibrium set of producers matched with those consultants, let $\mu(\mathcal{Z}^*)$ denote the measure of $\mathcal{Z}^*$, and let $m^* = \sup \mathcal{M}^* > 0$ denote the highest consultant type. We proceed in five steps. First, since consultant types are unobserved and consultant earnings are not contingent on their success, all consultants must earn the same expected payoff. Denote such payoff S^*. Second, note that $S^* \geq m$ for all consultants $m \in \mathcal{M}$ (as the outside option of a type m consultant is at least m) and therefore $S^* \geq m^*$. Third, as any agent may become a consultant, the equilibrium payoff of all producers is bounded below by S^*. Fourth, the combined (ex-ante) output produced by all matched agents ($\mathcal{M}^* \cup \mathcal{Z}^*$) is no larger than $\mu(\mathcal{Z}^*) \cdot m^*$ (as $\mu(\mathcal{Z}^*)$ is the mass of attempted problems and m^* is an upper bound for the probability that a given problem is solved). Fifth, since each producer must earn, in expectation, at least $S^* \geq m^*$, the combined earnings of all producers $\mathcal{Z}^*$ must be at least $\mu(\mathcal{Z}^*) \cdot m^*$. It follows that the combined earnings of all consultants $\mathcal{M}^*$ must be no greater than zero, a contradiction. ■

8 Appendix B: competitive equilibrium

In this Appendix we extend Remark 1 to the case in which the first best does not prescribe self-employment.

Remark 4 *Suppose the first best does not prescribe self-employment and let* $z^* = z_1^* = z_2^*$ *denote the marginal producer/consultant. In a full-information competitive equilibrium:*

A. $w_1^*(m) = C + \int_{z^*}^{m} \underbrace{n\left(Z^*(t)\right)}_{\frac{d}{dt}w_1^*(t)} dt$ *for all* $m \in \mathcal{M}$,

where $C = \frac{1}{1+n(z^*)}\left\{\Pi(1, z^*) - \int_{z^*}^{1} n\left(Z^*(t)\right) dt\right\} \geq z^*$.

B. $w_2^*(z) = M^*(z) - \frac{1}{n(z)} w_1^*(M^*(z))$ *for all* $z \in \mathcal{Z}$.

Proof. This proof is identical to the proof of Remark 1 except for the way in which the payment of the marginal consultant z^* is obtained. Such payoff, denoted C, is uniquely determined by the following two conditions. First, the total surplus available to a firm who hires $n(z^*)$ producers of type z^* and a consultant of type $M^*(z^*) = 1$ (namely, $n(z^*)$) must be fully allocated, per the zero-profit condition, to those agents:

$$w_1^*(1) + n(z^*) w_2^*(z^*) = n(z^*).$$

Second, payoffs must be continuous in type (i.e. $C = w_1^*(z^*) = w_2^*(z^*)$) and consultants must earn the full marginal value of their knowledge:

$$w_1^*(1) = C + \int_{z^*}^{1} n\left(Z^*(t)\right) dt.$$

Combing both equalities we obtain

$$n(z^*)\left(1 - C\right) = C + \int_{z^*}^{1} n\left(Z^*(t)\right) dt.$$

Note that the L.H.S. of the above equation is decreasing in C and the R.H.S. is increasing in C, and so we obtain a unique value for C. Moreover, that $C \geq z^*$ follows from the previous observation together with the fact that when $C = z^*$ the L.H.S. is no smaller than the R.H.S. per Lemma 2 (part 2). ■

REFERENCES

ACEMOGLU D., MOSTAGIR, M. and OZDAGLAR A. (2014), "Managing Innovation in a Crowd", (Mimeo).

AKERLOF, G. A. (1970), "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism", Quarterly Journal of Economics, 84 (3), 488–500.

ALCHIAN, A. and DEMSETZ, H. (1972), "Production, information costs, and economic organization", American Economic Review, 62 (5), 777-795.

DEMSKI, J. S. and SAPPINGTON, D. E. (1987), "Delegated Expertise", Journal of Accounting Research, 25 (1), 68-89.

EECKHOUT, J. and KIRCHER, P. (2012), "Assortative Matching with Large Firms: Span of Control over More versus Better Workers", (Mimeo).

ESWARAN, M. and KOTWAL, A. (1984), "The Moral Hazard of Budget-Breaking", Rand Journal of Economics, 15 (4), 578-581.

FUCHS, W. and GARICANO, L. (2010), "Matching Problems with Expertise in Firms and Markets", Journal of the European Economic Association, 8 (2-3), 354-364.

GALE, D. (2001), "Signalling in Markets with two-sided adverse selection", Economic Theory, 18 (2), 391-414.

GARICANO, L. (2000), "Hierarchies and the Organization of Knowledge in Production", Journal of Political Economy, 108 (5), 874-904.

GARICANO, L. and ROSSI-HANSBERG, E. (2004), "Inequality and the Organization of Knowledge", American Economic Review, 94 (2), 197-202.

GARICANO, L. and ROSSI-HANSBERG, E. (2006), "Organization and Inequality in a Knowledge Economy", Quarterly Journal of Economics, 121 (4), 1383-1436.

GARICANO, L. and SANTOS, T. (2004), "Referrals", American Economic Review, 94 (3), 499-525.

HAYEK, F. A. (1945), “The Use of Knowledge in Society”, American Economic Review, 35 (4), 519-530.

HERRENDORF, B., ROGERSON, R. and VALENTINYI, Á. (2009), “Two Perspectives on Preferences and Structural Transformation”, NBER Working Papers 15416.

INDERST, R. and OTTAVIANI, M. (2012), “Competition through commissions and kick Backs”, American Economic Review, 102 (2), 780-809.

KLEINER, M. M and KUDRLE, R. T. (2000), “Does Regulation Affect Economic Outcomes? The Case of Dentistry", Journal of Law and Economics, 43 (2), 547-582.

LELAND, H. E. (1979), “Quacks, Lemons, and Licensing: A Theory of Minimum Quality Standards", Journal of Political Economy, 87 (6), 1328-46.

LU, H. and ROBERT, J. (2001), “Optimal Trading Mechanisms with Ex Ante Unidentified Traders”, Journal of Economic Theory, 97 (1), 50-80.

MASKIN, E. and RILEY, J. (1984), “Monopoly with incomplete information”, Rand Journal of Economics, 15 (2), 171-196.

MILGROM, P. and SEGAL, I. (2002), “Envelope Theorems for Arbitrary Choice Sets”, Econometrica, 70 (2), 583-601.

MUSSA, M. and ROSEN, S. (1978), “Monopoly and product quality”, Journal of Economic Theory, 18 (2), 301–317.

MYERSON, R. and SATTERTHWAITE, M. (1983), “Efficient mechanisms for bilateral trading”, Journal of Economic Theory, 28 (2), 265-281.

OTTAVIANI, M. and SORENSEN, P. N. (2006), “Professional Advice”, Journal of Economic Theory, 126 (1), 120 -142.

PENSENDORFER, W. and WOLINSKY, A. (2003), “Second Opinions and Price Competition: Inefficiency in the Market for consultant Advice”, Review of Economic Studies 70 (2), 417 - 437.

SHAPIRO, C. (1986), “Investment, Moral Hazard and Occupational Licensing”, Review of Economic Studies, 53 (5), 843-862.

SORENSEN, M. (2007), "How Smart Is Smart Money? A Two-Sided Matching Model of Venture Capital", Journal of Finance, 62 (6), 2725-2762.

TAYLOR, C. (1995), "The Economics of Breakdowns, Checkups and Cures", Journal of Political Economy, 103 (1), 53-74.

WOLINSKY, A. (1993), "Competition in a Market for Informed consultants Services", Rand Journal of Economics, 24 (3), 380-398.