Employment as a Relational Obligation to Work

Michael Raith*

August 31, 2021

Abstract

This paper compares employment with spot trade of labor within the same economic environment. With spot trade, an entrepreneur and a wealth-constrained worker bargain over different actions, and without agreement there is no trade. Employment is a relational contract that gives the entrepreneur the authority to choose the worker's actions. Because of incomplete information about benefits and costs, either spot trade or employment may lead to inefficient outcomes. I derive a wide range of results about the entrepreneur's optimal choice between employment and spot trade. Employment is more efficient than spot trade the greater is the quasi-rent between the parties. Additionally, employment enables the entrepreneur to minimize the worker's rent. On the other hand, a wage premium is required to ensure the worker's obedience. My results suggest that the worker's *obligation* to provide his service is essential to the nature of employment. By contrast, whether the worker has multiple tasks, or the entrepreneur has an information advantage, or the entrepreneur owns important assets, all matter much less than the literature has suggested. A simplified version of the model also provides a simple theory of the firm that captures key ideas of transaction-cost economics.

*Michael Raith, Simon Business School, University of Rochester, Rochester, NY 14627, raith@simon.rochester.edu. I would like to thank Guido Friebel, Jin Li, Jeanine Miklos-Thal, Heikki Rantakari, Giorgio Zanarone, and seminar participants at the 14th Organizational Economics Workshop (ANU/UNSW), for helpful comments and suggestions.

1 Introduction

Why do firms hire employees instead of buying labor from independent contractors? What is employment anyway? First raised by Coase (1937), these questions are salient to 21st-century businesses. Large firms are increasingly replacing employees with contractors (Weber 2017, Bertrand et al. 2021). For small firms, opportunities to outsource services traditionally performed by employees abound, and even include high-level executive services (Anderson and Cappelli 2021, Broughton 2021). For workers, in turn, it is easier than ever to forgo traditional employment and be a freelancer or gig worker instead (Oyer 2016, Katz and Krueger 2019). Meanwhile, lawyers are debating how new work arrangements fit, or don't fit, within traditional definitions of employee and contractor (Posner 2020).

Coase and others argued that the exercise of authority over other people is a hallmark of firms, and employment a contract to establish an authority relationship (Coase 1937, Simon 1945, 1951, Arrow 1974).[1] [2] Authority and employment were believed to minimize the costs of adapting workers' actions to business needs (Coase 1937, Williamson 1975). Only limited progress has been made, however, on the key question of when and why authority—or more generally, fiat (Williamson 1971)—can outperform market transactions; Alchian and Demsetz (1972) even argued that the notion of authority itself is delusional. Informal discussions appeal to transaction costs, whereas formal comparisons of employment with market trade are scarce (see Section 2). This paper aims to help close this gap, and to shed new light on the nature of employment.

I consider a buyer and a seller of labor: an entrepreneur and a worker. The worker can perform one or more tasks for the entrepreneur, and can offer his services either as contractor or as employee. I abstract from effort choice, and asset ownership plays no direct role either. Instead, and in contrast to prior related work on employment, what drives my model is that the entrepreneur and the worker are privately informed about the benefits and (opportunity) costs, respectively, of the worker's actions.

As contractor, the worker bargains with the entrepreneur over prices of different tasks

[1] For instance, Simon (1945:177) wrote: "Of all the modes of influence, authority is the one that chiefly distinguishes the behavior of individuals as participants of organizations from their behavior outside such organizations." And Arrow (1974: 63): "the giving and taking of orders ... is an essential part of the mechanism by which organizations function."

[2] Throughout, by "authority" I mean "interpersonal authority" (Van den Steen 2010) without using the qualifier "interpersonal," consistent with its use by early organizational scholars and by sociologists (e.g., Coleman 1990). This differs from the recent broader use, within organizational economics, of "authority" as synonymous with managerial decision rights in general (Bolton and Dewatripont, 2013).

each time his services are needed. As employee, the worker confers to the entrepreneur the authority to choose the worker's action, in return for a wage. This notion of authority follows Coase and Simon (1951) and is purely transactional: the worker is expected to obey the entrepreneur for no other reason than his own prior agreement to obey.[3] Unlike in Simon's theory, however, employment is distinct from spot trade even if there is only one task: The contractor can always decline to perform a service, whereas the employee is obligated to work, and in my theory this emerges as the key definitional distinction between the two, irrespective of job design.

The employee's obligation to work is relational, not legal, in my model as it is in many legal environments. Although the worker transfers his control right voluntarily, the agreement is not enforceable, and dismissal is the entrepreneur's only sanction for disobedience.[4] The worker therefore obeys in order to keep his job, which is worthwhile for him if the wage is high enough, and worthwhile for the entrepreneur if not too high.

Envisioning employment as a relational obligation reconciles Coase's and Simon's (1951) contractual approach with other authors' emphasis on the limits of authority (Barnard 1938, Simon 1945, Arrow 1974). At the same time, it rebuts Alchian and Demsetz's argument against the notion of authority: unlike a contractor who refuses to trade, an employee who disobeys is in breach of an agreement he entered into, and stands to lose the benefits of the relationship. In my model, the sole purpose of a relational contract is to support the entrepreneur's authority over the worker.

Private information about the benefits and costs of the worker's actions endogenously leads to inefficiency, or "transaction costs," both with spot trade and with employment. If the worker is a contractor, the parties may fail to agree on a price even if trade is efficient, as is well known from the bargaining literature (Myerson and Satterthwaite 1983, Ausubel et al. 2002). But employment may be inefficient too if the entrepreneur orders an action whose cost to the worker exceeds the benefit. Even if the parties can bargain following

[3]See also Coleman (1990, p.73): "In [a formal organization composed of employees working for pay], transfer of the right to control is made in exchange for payment of a wage or salary. There is no assumption that authority will be exercised in the interest of the actors ... who have transferred the right"

[4]Barnard (1938: 165) argued, "A person can and will accept a communication as authoritative only when ... at the time of his decision, he believes it to be compatible with his personal interest as whole." Consequently (Arrow 1974), "Within the firm, the sanctions which authority can use are basically those of hiring and firing" and "The scope of this authority ... is limited by the freedom with which an employee can leave the job." Enforceability depends on the legal environment, though. In Britain, "until 1875, when it was repealed, Master and Servant law gave employers the ability to criminally ... prosecute and severely punish a majority of employees across industries for breach of contract" (Naidu and Yuchtman 2013, see also Pollard 1963, Hay and Craven 1993).

such an order—for example, the employee may want to negotiate time off for an important personal errand, to be made up another time—incomplete information may still stand in the way of reaching efficiency.[5]

The framework described thus far (and formally introduced in Section 3) permits comparing spot trade with employment within the same economic environment (see Hart 1995, Gibbons 2005). It highlights the key difference between market and hierarchy: With spot trade, no trade is the default; with employment, the default is the action ordered by the entrepreneur. The difference matters because costly bargaining is naturally biased towards the respective disagreement point (Farrell 1987, McKelvey and Page 2002). Observed benefits and costs of each structure—notably, the dichotomy "fiat vs. haggling"—all result endogenously from the underlying control rights.[6]

It follows as first main result (see Sections 4 and 5) that employment is more efficient relative to spot trade if and only if, broadly speaking, the surplus at stake, i.e., the quasi-rent, is large enough. The larger the surplus, the more efficient on average will it be for the worker to work on some task than not to work, which is what employment establishes as default. This result echoes Coase's (1960) argument that property rights matter for outcomes if and only if bargaining is costly (see also Coase 1992, Tadelis and Williamson 2013, Section 2.1).

In another departure from related work, I assume that the worker is wealth-constrained, reflecting the reality faced by workers throughout much of history. The worker is therefore unable to pay the entrepreneur upfront for any rents earned either as contractor or as employee. It follows that the entrepreneur's choice between employment and spot trade may not be driven by efficiency considerations alone, but by rent extraction motives too.[7]

To study how profit maximization may differ from efficiency, I assume that wealth-constrained workers compete ex ante to do business with the entrepreneur. Similar

[5] A lack of court enforcement (Williamson 1991) does not prevent the parties from reaching spot agreements within employment, because compliance is supported by the same relational contract that supports the relationship to begin with.

[6] In my model, a relational contract at arm's length cannot improve upon spot trade because valuations are private whereas actions are contractible; see Section 3.3. Employment and sequential spot trade are the two main contenders in Williamson's (1975: 64-72) discussion of "individualistic bargaining modes". Williamson dismisses ex-ante contracting, Simon's (1951) alternative to employment, as "singularly unsuited to permit adaptation in response to changing ... circumstances" (p. 65). In turn, "contingent claims contracting ... fails principally because of bounded rationality" (p. 72).

[7] Claims that hierarchical production and wage labor are exploitative go back to Marx (1867), see also Marglin (1974). For further discussion, see Section 9, Williamson (1985) and Rebitzer (1993).

as in Hart and Moore (2008), once the entrepreneur chooses a worker as well as her profit-maximizing governance structure (employment or spot trade), a Williamsonian "fundamental transformation" (1985: 61-63; 1975: 61-64) creates appropriable quasi-rents and thus in effect a bilateral monopoly between entrepreneur and worker. These quasi-rents could result from "purposeful investments" or could just be due to "knowledge and skills that are incidentally acquired by the parties while working together" (Tadelis and Williamson 2013).

The first few results explained below do not rely on any assumptions about bargaining protocols. For the others, I assume that the party that does *not* hold the control right over the worker's action—with spot trade, the entrepreneur; with employment, the worker—makes a take-it-or-leave-it price offer, which the control-right holder accepts or rejects. This assumption, which builds on Bajari and Tadelis (2001), makes bargaining with two-sided incomplete information tractable even in more complex settings, and is consistent with rent-seeking theories which posit that people will expend resources to influence those holding decision rights. As importantly, it also maximally biases the entrepreneur's profits in favor of spot trade and thus raises the burden of proof for the Coasian claim that authority can outperform the market. I derive the following results:

1. Employment may be less profitable for the entrepreneur even if it is more efficient than spot trade (Section 3). This follows because relational authority requires an "obedience wage premium", akin to an efficiency wage markup, for the worker's continuation utility to exceed any gains from disobedience. If that wage premium is too high, and the worker is unable to "buy" the job upfront, the entrepreneur may prefer spot trade.

2. Employment may be more profitable even if spot trade is more efficient. That is because spot trade typically leaves the worker an information rent for which he cannot compensate the entrepreneur upfront. With employment, in contrast, the entrepreneur can hold the wage close to the worker's expected cost if the parties are sufficiently patient. That is, ex ante, workers vying to be contractors may be underbid by workers willing to work as employees for a low wage, even if employment is inefficient.

3. Employment is more profitable the less the parties discount future payoffs, or equivalently, the more frequently they interact. The lower the discount rate, the lower is the obedience wage premium, and the more of the worker's information rent can be extracted.

4. Result 2 implies that contrary to an idea that runs through most of the literature on

employment, the entrepreneur need not be better informed about what the worker should be doing, for employment to be profitable (Section 4). She may choose employment even if her valuation for work is constant and known, as long as it is high enough. By contrast, variability in the worker's costs affects the entrepreneur's profits both with spot trade and with employment, and gives employment the edge when the parties are patient enough.

5. Also in contrast to the entire literature, a multiplicity of tasks and a boss's ability to tell the employee *what* to do are inessential to the nature of employment in my model (Section 5.2). Both the employed and the independent IT specialist will manage my database and not my website, if that is what I currently need. The real difference between the two is that the contractor can decline my request for service, or demand a surcharge, whereas the employee is relationally obligated to work for me at an agreed price. This difference, however, is already present with only one task, which indirectly helps explain the ubiquity throughout history of wage labor with extreme specialization.

6. Bundling different tasks into one job can significantly raise the profitability of employment if the firm's benefits are negatively correlated across tasks, and the worker's costs positively (Section 5.3). While the entrepreneur may not have enough demand to hire specialists for database and website management, she may well have demand for one IT person to work on these tasks at different times, especially if the worker cares little about which task is requested. In practice, then, employees often do have multiple tasks because firms will pay a regular wage only to people who always have something to do.

7. It is possible but only in special cases optimal for the entrepreneur to delegate the choice of tasks to the worker (Section 6). Employment with delegation differs from spot contracting in that the employee is still obligated to provide a service, whereas the contractor is not. Delegation may be advantageous if the costs of tasks are more variable than their benefits. The "empowered" worker will settle for a lower wage, but a lower wage tightens his obedience constraint, which in turn reduces the entrepreneur's profit. I show that delegation may be optimal for the entrepreneur if spot adjustments are infeasible, but is never optimal with unrestricted spot adjustments.

Results 5 and 7 suggest that *defining* employment as a worker's obligation to work may be more useful than defining it by the firm's authority over the worker's actions. This broader definition both covers jobs with only one task, and it permits, for jobs with multiple tasks, a clear distinction between employment with delegation and independent contracting. Result 7 helps explain, however, why with employment the labor buyer is

also typically the "boss."

8. Reasons to limit the entrepreneur's formal authority, previously analyzed only by Simon (1951), are less obvious than one might expect (Section 7). First, if some actions are costlier than others, it may be better to tailor the wage to the task than to exclude it, as is common practice with overtime pay. Second and more surprisingly, I show that under weak conditions, it is always preferable to include rather than exclude a task so long as it is *sometimes* efficient. What drives this result is that the entrepreneur's order conveys information about her valuations, which makes it easier to negotiate away from inefficient orders than to negotiate towards efficient actions outside of the acceptance set. In practice, then, the reason to limit authority may simply be to circumscribe the worker's obligations and to determine his wage.

9. With multiple workers, complementarity in production favors employment (Section 8). Spot trade is less profitable in that case because the entrepreneur pays more to secure all workers' participation. Meanwhile, employment may be more profitable if forcing some workers to work, even when their cost is high, is efficient for team production. This result helps to explain why wage labor has historically correlated with large-scale production, but may become less dominant as firms adopt more flexible production methods.

10. If a Williamsonian fundamental transformation is absent and workers compete for tasks *ex post*, then with enough workers, spot trade unambiguously dominates employment because the entrepreneur can offer a low price that is likely to be accepted by at least one worker. This is the business model of companies such as ridesharing service Uber or labor marketplace TaskRabbit.

Section 9 argues that clues supporting this paper's assumptions and results can be gleaned from both the 18th century and the 21st century, which, respectively, mark periods of growth and decline of wage labor. Section 10 relates the theory to the current legal discussion about the employee-contractor distinction.

In Section 11 I argue that, moving beyond employment, the core model of bilateral trade provides a simple but fully micro-founded theory of the firm that captures key ideas of transaction-cost economics, including its central prediction that vertical integration is more likely the greater the appropriable quasi-rents. The model blends what Gibbons (2005) has called the adaptation and rent-seeking approaches to the theory of the firm, as two sides of the same coin. Section 12 concludes.

2 Related literature

Although economists from Knight to Williamson wrote extensively about authority and employment, formal theoretical analyses are scarce. Comparing authority with spot trade requires a model in which ex-post agreements are costly but feasible. If instead they are costless as much of property-rights theory assumes, then spot trade is already efficient and authority has no role. But if spot agreements are infeasible as another large literature assumes, then authority and market cannot be compared.

Methodologically most closely related to my paper is Bolton and Rajan (2003), see also Bolton and Dewatripont (2004: 588, and 2013: 369). Employment is modeled as authority supported by a relational contract, and incomplete information may lead to inefficiency. However, Bolton and Rajan assume that the buyer knows the full state of nature, including the seller's cost. The purpose of employment is then to utilize the buyer's superior information, and employment is always efficient if it is feasible. My informational assumptions are both more general and more conventional, and employment is not necessarily more efficient or more profitable than spot trade.

Some papers compare employment with independent contracting, while abstracting from obedience as a problem. Hart and Moore (2008) compare employment with long–term independent contracting and with spot trade, in a framework in which the costs of ex-post contracting are derived from behavioral assumptions. Wernerfelt (1997, 2004, 2015) focuses on optimal adaptation in the presence of bargaining and other adjustment costs, which he models in reduced form. Levin and Tadelis (2010) compare contracting on time (employment) with contracting on quality (outsourcing), under the assumption that time is easy but quality is costly to measure. In contrast to these papers, the costs of contracting in my model result endogenously from incomplete information.

Other papers focus on obedience but do not allow for market trade. In Van den Steen's (2010) theory of the firm, concentrated asset ownership, authority, and low-powered incentives go hand in hand. Authority is needed for two parties to coordinate on how to run a project when they disagree due to differing priors. Van den Steen's model does not permit trade, however. In equilibrium, two entrepreneurs either choose to be integrated, or they remain independent and make their own decisions.

In Rantakari (2021), authority can emerge as a relational contract between two symmetric players involved in a coordination problem. The source of the superior's authority

is an information advantage gained through endogenous discovery effort, which relates to Barnard's (1938) discussion of the legitimacy of authority. In Rantakari's model, too, trade is infeasible; the main alternative to an authority relationship is relational coordination between independent, symmetric agents.

Marino et al. (2010) focus on the constraints that potential disobedience imposes on the allocation of decision rights in firms. They show that employees with attractive outside options, or who are hard to replace, may need to be given more authority in order to prevent disobedience.

A third strand of literature focuses on how the ex-ante allocation of control rights facilitates adaptation, in contexts other than employment. Like mine, several papers in this literature feature ex-post bargaining under incomplete information. Bajari and Tadelis (2001) study how procurement contracts solve a tradeoff between providing incentives and reducing ex-post transaction costs. Section 11 relates my model to Bajari and Tadelis's when both are interpreted as theories of the firm. Matouschek (2004), Segal and Whinston (2016), and Baliga and Sjöström (2018) develop property-rights theories of the firm based on the objective to minimize the inefficiency of ex-post bargaining; see also Segal and Whinston (2013). Chakravarty and MacLeod (2009) show how authority provisions in construction contracts regulate the parties' ex-post bargaining power, in the shadow of legal constraints on enforcement. Baker, Gibbons, and Murphy (2011) study how decision rights support adaptation through the use of relational contracts when the state of the world is observable to both parties. Their Section 4 allows for ex-post contracting with exogenous bargaining costs. Zanarone (2013) shows how franchise contracts facilitate adaptation in the shadow of laws designed to protect franchisees. Powell (2015) compares integrated and non-integrated structures when the goal is to facilitate adaptation and to curtail wasteful influence activities. Like in Van den Steen (2010) and in Rantakari (2021), Powell's setting is a coordination game without market trade.

In spite of overlap with this literature, three features of my model stand out: First and most importantly, the default outcome under one structure (spot trade) is no trade, i.e., no adaptation, whereas under the other (employment) it is unilateral adaptation, which follows naturally from the benefits and costs of different actions. This asymmetry explains why ex-post bargaining is more important in the market than in the firm; i.e. "fiat vs. haggling" are endogenous outcomes, not a definitional distinction. Second and unique to employment, the entrepreneur's control rights over the worker's actions can only be

relational; that is, legal control rights always rest with the worker. Third and relevant for some of my results, wealth constraints and asymmetry in the labor market can lead to inefficient equilibrium governance structures, in contrast to a focus on efficiency in much of organizational economics.

Related to my key question but methodologically quite different are papers in the property-rights and incentive-system literatures that define employment by a firm's ownership of the assets that are used by the worker, such as Grossman and Hart (1986), Holmstrom and Milgrom (1991, 1994), Baker, Gibbons, and Murphy (2002), and Zanarone (2012). This perspective is complementary in that the literature makes explicit how asset ownership shapes outside options and therefore feasible equilibria. However, it focuses on incentives; decisions and authority play no role.[8] In Section 9 I argue that it is both methodologically and empirically unsatisfactory to *define* employment in terms of asset ownership. In my model, assets and their ownership need not play any particular role for employment to be well defined and to be distinct from spot trade.

Accordingly, property-rights and incentive-system theory and my paper have different responses to Alchian and Demsetz's (1972) challenge to the Coase-Simon notion of authority. Property-rights theory sides with Alchian and Demsetz's claim that there is no *conceptual* difference between firing a grocer and firing an employee, but argues that the ownership of non-human assets confers power over people: the fired grocer still owns his assets and therefore has other customers to sell to; the fired employee may have much more limited options (Hart 1996, Roberts 2004: 104, Van den Steen 2010: 477-478). In my model, by contrast, authority means the same as it has meant to sociologists since Max Weber: one agent's acquiescence to the direction of another (Simon 1945:10; see also Blau 1964: 200 and Coleman 1990: 67). Specifically, the worker as employee voluntarily and explicitly vests authority over his actions in the entrepreneur (Coleman, chapter 4). Legal enforceability is not needed if the contract can be self-enforcing, but the distinction is still clear: the employee who refuses to "type this letter" is in breach of an agreement, the grocer who refuses to "sell me this brand of tuna" is not.[9]

Finally, within a large and growing recent literature on relational employment con-

[8] An exception is Zanarone (2012), who points out that contracts involving decisions by one party's fiat can also occur between firms, such as in franchise systems. In his model, asset ownership defines integration or non-integration, whereas fiat (within or between firms) is supported by a relational contract.

[9] This argument also differs from Williamson's (1991) assertion that a firm's authority originates from the legal doctine of forebearance; see Section 10 for further discussion.

tracts, Li and Matouschek (2013) and Li et al. (2020) are closest in that they, too, assume that the worker has limited liability and the principal has private information. However, this literature focuses on the dynamics of employment relationships and is concerned with very different questions.

3 General framework

3.1 Model

An entrepreneur E and a worker W are engaged in a long-term relationship. As in Hart and Moore (2008), ex ante there are many workers competing to supply their labor. Once a worker is chosen, however, competition is significantly reduced due to a "fundamental transformation" in the sense of Williamson (1985). That is, relationship-specific investments or the incidental acquisition of knowledge and skills (Tadelis and Williamson 2013) may in effect lead to a bilateral monopoly—the creation of appropriable quasi-rents—between E and W. Like in Hart and Moore, all that matters is the presence of a fundamental transformation, not its precise causes.[10] Exclusivity is not assumed: E may have multiple workers, and W may work for multiple firms, even as employee (e.g., in two part-time jobs).

Unlike in Hart and Moore, however, the workers have no wealth to buy themselves into a relationship with E; that is, they are unable to pay E upfront for any rents they might earn either as contractor or as employee. This assumption formalizes Williamson's (1975: 69-70) detailed discussion of why workers may be unable to pay upfront for any rents they would earn as contractors; notably, "it is seriously to be doubted that they could raise the funds, if their personal assets were deficient, to make the implied full valuation bids."

Once matched in $\tau = 0$, E and W engage in trade over potentially infinitely many periods, starting in $\tau = 1$. If they fail to trade in any period, each earns 0 in that period. If they break up, then going forward, both E and W earn 0 in every period. Equating the outside option payoffs with the no-trade payoff means that the benefits and costs of trade, described below, are to be understood as opportunity benefits and costs, relative to outside options. In particular, fluctuations in W's costs may be primarily due to changing

[10] The same process towards "small-numbers exchange" is described in more detail in Williamson (1975: 61-64). The fact that this transformation often occurs organically and without purposeful investments justifies ignoring any such investments in the analysis; see Gibbons (2010) for a discussion of this point.

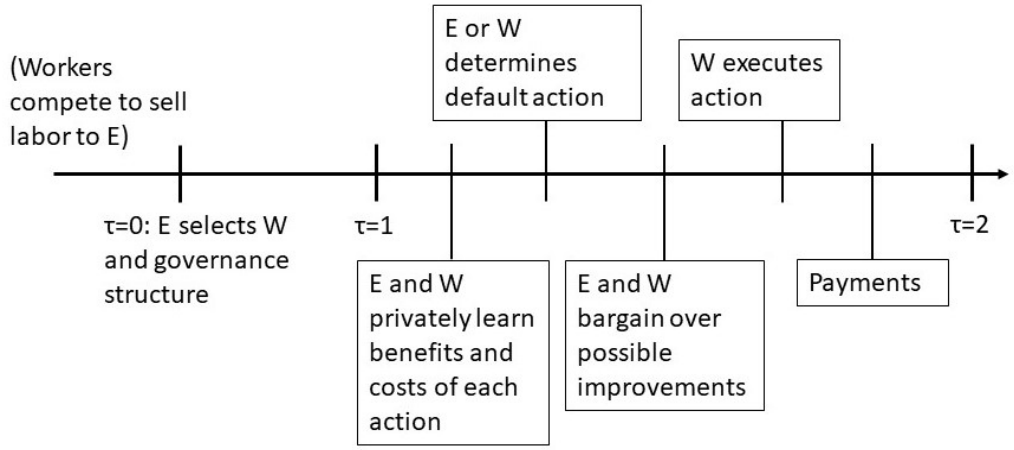


Figure 1: General timeline of events

outside options rather than changing costs or preferences of working for E.

Both players are risk neutral and discount future payoffs at a common discount factor δ. Each player maximizes, at each point in time, their net present value of payoffs. To abstract from savings dynamics, I assume that W has no access to credit markets. A timeline of events is sketched in Figure 1.

In each period, W can carry out one of M possible actions (or tasks) indexed by $1...M$, or alternatively not work for E, which is action 0 ("no work" is a shorthand for pursuing the next best use of time). Each task $k \geq 1$ generates benefit $b^k \in B$ to E and causes cost $c^k \in C$ to W, for finite B and C with strictly positive elements. If W does not work, then $c = b = 0$.

Denote the players' type spaces by $\mathcal{B} = B^M$ and $\mathcal{C} = C^M$. Both players have common prior beliefs that $b = (b^1, ..b^M)$ is distributed according to the probability distribution $p : \mathcal{B} \rightarrow [0,1]^M$, and $c = (c^1, ..c^M)$ according to $q : \mathcal{C} \rightarrow [0,1]^M$. I assume that p and q are independent and are i.i.d. across time periods. They can be correlated across tasks but have identical marginal distributions for each task. If W carries out task k, the surplus generated between E and W is $s^k = b^k - c^k$.

All tasks are contractible, although with employment E and W do not actually write contracts over actions. For most of the paper, I consider two different governance structures. One is spot trade, in which E and W bargain and trade in each period independently, with no work being the default (disagreement) outcome. The other is employment, which is a relational contract that confers to E the right to control W's actions out of a

subset $A \subseteq \{1, .., M\}$, in return for a constant wage w (I comment below on the optimality of a constant wage). Disobedience is followed by severance of the relationship; i.e., W cannot switch to spot trade.[11] As a result of ex-ante competition among workers, E chooses, along with the identity of W, the governance structure that leads to the highest payoff or her at $\tau = 0$.[12]

For both spot trade and employment (for which I use superscripts 0 and E below), the timeline within each period $\tau \geq 1$ is the same:

1. E privately learns $b = (b^1, ..b^M)$ and W privately learns $c = (c^1, ..c^M)$.

2. The party with the decision right over W's actions determines their optimal default action. For W as contractor, that is not to work; denote this choice by

$$\widetilde{m}^0 = \arg \min_{k \in \{0,..M\}} c^k = 0.$$

   With employment, E's optimal action is $\arg\max_{k \in A}\{b^k\}$. In case of a tie, E chooses the action that maximizes the surplus s^k. Although E does not know c, this choice can be thought of as outcome of an incentive-compatible direct mechanism that selects for Pareto-efficient outcomes subject to maximizing E's payoff. . Thus, given $b \in \mathcal{B}$ and $c \in \mathcal{C}$, denote E's optimal default action by

$$\widetilde{m}^E(b, c) = \arg\max_l\{s^l | l \in \arg\max_{k \in A}\{b^k\}\}, \tag{1}$$

   which E communicates to W as "order".

3. E and W engage in bargaining over all actions $\{0, ..M\}$ (including no work), according to a direct mechanism (m^i, t^i) for $i \in \{0, E\}$, whereby E and W agree on action

[11] This assumption is consistent with Levin (2003: 840) but is in contrast to Baker et al. (2002: 50), who assume that the parties revert to spot trade if their relational contract terminates. The threat of severance is credible if even after $\tau = 0$ there continue to be other workers available to take W's spot, and if the quasi-rents between E and W (i.e., the costs of turnover) are not so large as to shut out other workers as alternative sellers. It follows, importantly, that W's payoff as employee can be lower than his payoff as contractor, if that is what maximizes E's profit. An employee who were to say, "From now on I wish to be a contractor and charge you higher average rates" would simply be fired. In contrast, Baker et al. (2002) assume reversion to spot trade because there are only two parties.

[12] Once in place, there is no scope to renegotiate the wage: although both sides face exit costs, E has no reason to agree to any wage increase, and can credibly fire a worker who demands a raise just as she can fire a worker who shirks. As Malcomson (1999, Section 4.2) argues, even if the firm cannot legally dictate the wage, "employment at will as interpreted by US courts has exactly the same effect."

$m^i(b,c)$ in return for a transfer $t^i(b,c) \geq 0$ from E to W in the case of spot trade, and from W to E with employment.[13] The mechanism selects for Pareto-efficient outcomes subject to incentive compatibility and can depend on i in three ways. First, different bargaining protocols may be used for spot trade and employment. Second, the mechanism must satisfy interim rationality constraints defined by the default action $\widetilde{m}^i(b,c)$. For spot trade, that requires

$$b^{m^0(v,c)} - t^0(v,c) \geq 0 \quad \text{and} \quad t^0(v,c) - c^{m^0(v,c)} \geq 0,$$

for E and W, respectively, and for employment,

$$b^{m^E(v,c)} + t^E(v,c) \geq b^{\widetilde{m}^i(v,c)} \quad \text{and} \quad c^{m^E(v,c)} + t^E(v,c) \leq c^{\widetilde{m}^E(v,c)}$$

(see Segal and Whinston, 2016). Third, lacking access to credit markets, W as employee cannot pledge more than his current wage as transfer to E, therefore $t^E(v,c) \leq w$. Because disagreement is a possible outcome, there will usually be types (b,c) for which $m^i(b,c) = \widetilde{m}^i(b,c)$ and $t^i(b,c) = 0$.

4. With spot trade, W executes $m^0(b,c)$; i.e., the task that E and W agreed on, or no work. With employment, if W executes $m^E(b,c)$, the relationship continues into the next period. If W disobeys by choosing $l \neq m^E(b,c)$, then the relationship terminates immediately.

5. Payments are made, as applicable. With spot trade, E pays $t^0(v,c)$ as per agreement. With employment, E's payment to W is discretionary. If E pays $w - t^E(b,c)$, then the relationship continues into the next period. If E pays any other amount or nothing, then the relationship terminates immediately.

Between periods, and between all stages within each period, E and W can unilaterally terminate their relationship if they wish, consistent with an "employment at will" legal doctrine (see Malcomson 1999 for a discussion). Realistically, in modern legal regimes employment is enforceable at least to the degree that E legally owes W his wage in any

[13] For now, the stated non-negativity of transfers does not restrict the bargaining mechanisms, but simply follows from the fact that if the decision-right holder chooses the unilaterally optimal default action, any mutually beneficial adjustment must entail a nonnegative transfer from the other party.

period in which W works for E. However, to keep the notion of relational authority clean, I assume that no such legal enforcement exists.

When negotiating spot adjustments, the unenforceability of any agreements poses no problem because E and W can piggyback on their relational contract. That is, since authority needs to be self-enforcing even without adjustments, and because any adjustment is a Pareto improvement, there is no reason for either party to renege on a spot adjustment if the basic enforcement constraints (derived below) already hold.

With spot trade, the default benefit and cost are $b^0 = c^0 = 0$. With employment, if E's order $\widetilde{m}^E(b,c)$ is carried out without further adjustment, the expected benefits and costs are

$$b^E = E_{b,c}\left[b^{\widetilde{m}^E(b,c)}\right] = \Sigma_b \Sigma_c p(b)q(c)b^{\widetilde{m}^E(b,c)} \text{ and } c^E = E_{b,c}\left[c^{\widetilde{m}^E(b,c)}\right].$$

Next, E's and W's expected net gains from spot adjustments are, for spot trade,

$$V^0 = E_{b,c}\left[b^{m^0(b,c)} - t^0(b,c)\right] \text{ and } U^0 = E_{b,c}\left[t^0(b,c) - c^{m^0(b,c)}\right]$$

and for employment,

$$V^E = E_{b,c}\left[b^{m^E(b,c)} - t^E(b,c) - b^{\widetilde{m}^E(b,c)}\right] \text{ and } U^E = E_{b,c}\left[c^{\widetilde{m}^E(b,c)} - t^E(b,c) - c^{m^E(b,c)}\right].$$

I have assumed, like Simon (1951), that employment pays a constant wage w in every period. In fact, such a contract is optimal even if the wage can be task- or time-contingent, as long as the wage profile does not affect W's ability to negotiate spot adjustments. To see this, suppose that more generally, a relational employment contract defines for each period $\tau \geq 1$ and each history of the game h^τ leading up to τ a wage function $w(h^\tau) : A \to \mathbb{R}$.

Proposition 1. *For any wage function $w(h^\tau)$ there exists a stationary wage $w \in \mathbb{R}$ that satisfies all incentive constraints and leads to the same sequence of orders $\{\widetilde{m}^E_\tau\}$. If in addition for an optimal contract $w(h^\tau)$ spot adjustments are either infeasible, or are feasible and the corresponding stationary wage satisfies*

$$w \geq \max\{b \in B | \exists c \in C \text{ such that } b < c\}, \tag{2}$$

then a stationary contract is optimal.

There are two parts to this result: First, the optimal wage does not depend on the task because under the independence and symmetry assumptions about p and q, a constant wage ensures that E's profit-maximizing action is also efficient, contingent on E's information, and leads to the least restrictive incentive constraints for W. Second, in spite of W's limited liability, the optimal contract is stationary because (in this model) employment is not an incentive contract. In incentive models with limited liability, backloading payments to the agent is often optimal because it strengthens incentives in future periods (see, for instance, Barron et al., 2018). Here, backloading wages would relax future incentive constraints too, but E cannot gain from that because she already picks the action that is optimal for her in each period. As a result, all incentive-compatible wage profiles lead to the same stochastic sequence of E's orders, and it is weakly optimal for the parties to settle up in each period.

The wage profile can affect actions, however, via W's financing constraint $t^E(v,c) \leq w$ for spot adjustments. If, for instance, a stationary wage is too low for W to be able to negotiate away from an inefficient order, an improvement could possibly be reached by paying less in one period and more in another.

3.2 Enforcement constraints and E's profit

Focusing on stationary contracts, E's and W's expected employment payoffs in each period are

$$\widetilde{V}^E = b^E - w + V^E \quad \text{and} \quad \widetilde{U}^E = w - c^E + U^E, \tag{3}$$

where, per the definitions above, b^E and c^E are the expected benefit and cost of E's orders and V^E and U^E are the expected gains for E and W from any spot adjustments. If, for whatever reasons, spot adjustments are infeasible, all derivations below go through with $V^E = U^E = 0$. Let $b_L = \min B$ and $c_H = \max C$. By symmetry of p, any action $l \in A$ may be optimal for E, and by symmetry of q and independence of p and q, c_H is a possible realization for any $l \in A$. Therefore, E will order with positive probability an action for which W's cost is c_H. W can quit in this case, and receive a zero payoff now and in the future. The most restrictive condition that ensures S's obedience therefore is

$$w - c_H + \sum_{\tau=1}^{\infty} \delta^\tau \widetilde{U}^E \geq 0$$

or simpler,

$$\text{(WIC)} \quad (1-\delta)(w-c_H)+\delta\widetilde{U}^E \geq 0. \tag{4}$$

E, in turn, could fire W if her benefit is $b_L < w$, but more tempting, if legally feasible, would be to renege on the promised wage after W has executed E's order. The most restrictive constraint to prevent that is

$$\text{(EIC)} \quad (1-\delta)(-w)+\delta\widetilde{V}^E \geq 0.^{14}$$

Denoting the total surplus from employment by $S^E = \widetilde{U}^E + \widetilde{V}^E$, and combining (WIC) and (EIC) to a joint enforcement constraint, employment as relational contract is feasible if $-(1-\delta)c_H + \delta S^E \geq 0$, or

$$\delta \geq \delta_0 = \frac{c_H}{S^E + c_H}. \tag{5}$$

Thus, not surprisingly, employment is feasible only if the discount factor exceeds a lower bound. In what follows, however, I will be more concerned with profitability than with feasibility. That is, I will focus on δ for which (5) is slack, and examine the forces that determine E's choice between spot trade and employment. Given feasibility, the profit-maximizing employment contract pays the lowest possible wage w determined by (WIC).

It will be convenient to express E's profits in relation to the first-best surplus. Denote the efficient action for each b and c, which may equal no work, by $m^*(b,c) = \arg\max_{l\in\{0,\dots M\}}\{s^l\}$.Then the expected first-best surplus is $S^* = E\left[s^{m^*(b,c)}\right]$. Define the expected deadweight loss D^i as the difference between S^* and the expected surplus from spot trade or employment via the identity

$$S^* = b^i - c^i + U^i + V^i + D^i, \text{ for } i = 0, E, \tag{6}$$

where $b^i - c^i$ is the expected surplus from the exercise of i's authority without adjustments (with spot trade, that is zero), and $U^i + V^i$ is the expected gain in surplus from spot adjustments, with all variables defined in Section 3.1. Then with spot trade, E's profit is

$$V^0 = S^* - U^0 - D^0. \tag{7}$$

[14] See the discussion if Section 3.1 immediately following stage 5 of the timeline. If (depending on the legal environment) E were required to pay a wage after W has worked, the constraint would be $(1-\delta)(b_L - w) + \delta\widetilde{V}^E \geq 0$. This would affect the minimum discount rate needed for employment to be feasible, but would not affect any of the paper's main results.

That is, E's profit is the first-best surplus, minus rents that the worker earns, and minus any remaining deadweight loss. With employment, w is determined by (WIC) and (3):

$$w = c^E + (1-\delta)(c_H - c^E) - \delta U^E. \tag{8}$$

The profit-maximizing wage consists of three parts: First, W needs to be compensated for the costs associated with E's orders. Next, a wage premium $(1-\delta)(c_H - c^E)$ is required in order to satisfy W's obedience constraint (4). Because (4) is in effect a no-shirking condition, this premium resembles an efficiency wage premium (see also Van den Steen 2010). Finally, depending on the magnitude of δ, E can partially recoup W's adjustment gain in the form of a lower wage, as compensating differential. Plug (8) for w into E's profit, and then replace $b^E - c^E$ using (6), to obtain

$$\widetilde{V}^E = W^* - D^E - (1-\delta)(c_H - c^E + U^E), \tag{9}$$

where the last term is W's payoff, $\widetilde{U}^E = (1-\delta)(c_H - c^E + U^E)$. Thus, with employment, E can capture the first-best surplus, minus any remaining deadweight loss after spot adjustments, minus two kinds of rents earned by W that disappear as δ nears 1: the obedience wage premium $(1-\delta)(c_H - c^E)$, and W's gains from spot adjustments that E can partially recoup through a lower wage. Combining (7) and (9), we obtain

Proposition 2. *The difference between E's profits from employment and from spot trade is given by*

$$\begin{aligned} \Delta V &= \widetilde{V}^E - V^0 = D^0 - D^E + U^0 - (1-\delta)(c_H - c^E + U^E) \\ &= D^0 - D^E - (1-\delta)(c_H - c^E) + U^0 - (1-\delta)U^E, \end{aligned} \tag{10}$$

which is increasing in δ.

Proposition 2 shows that the profitability of employment relative to spot trade is driven by three effects. To begin with, employment is more profitable the more efficient it is relative to spot trade, as reflected by the term $D^0 - D^E$. Without imposing more structure on the model, however, nothing more can be said about this formally, and I postpone a discussion until Section 4.

Second, the profit is reduced by the obedience wage premium $(1-\delta)(c_H - c^E)$. As a

result, E may prefer spot trade even if employment is more efficient, see Section 4 for an example. Thus, employment requires not only a minimal discount factor to be feasible, but requires a higher discount factor to be profitable for E.

Third, the expression $U^0 - (1-\delta)U^E$ in (10) shows that W's rents under spot trade and employment affect E's profit differently: while any rents earned by W with spot trade are lost to E, with employment, she can partially recoup W's adjustment gains through a lower wage, depending on δ. Like the second effect, this difference for E's profit results from W's wealth constraint—without, E could capture, under either arrangement, all of W's rent through an upfront payment. E's loss of U^0 with spot trade formalizes two connected arguments made by Williamson (1975: 67-70): first, that spot trade is "impaired by ... problems of opportunism" (p. 72), and second, that workers are unlikely to be able to "submit lump sum bids for jobs at the outset" (p. 69).

The profit difference ΔV increases with δ, leading to the prediction that the more frequently E and W interact, the more likely E prefers employment. Not only is employment more likely to be feasible according to (5), but E's ability to capture surplus as profit increases with δ as well, because the obedience premium is lower and because E can recoup a larger share of W's adjustment benefit U^E the more W values future payoffs. This prediction is supported by Abraham and Taylor's (1996) evidence that firms outsource labor to smoothen out fluctuations in demand for labor or when they need specialists but only infrequently.

As $\delta \to 1$, ΔV reduces to $D^0 - D^E + U^S$. The entrepreneur then prefers employment not only if and because it is more efficient than spot trade, but because employment leaves no rent to W whereas spot trade normally does. As I show in the next section, because of this effect, employment may also be most profitable even if spot trade is more efficient.

3.3 Relational outsourcing?

Given that employment as defined above is a very specific relational contract, it is natural to ask what other relational contracts might exist. Short of providing a formal result, I will argue that there is no scope for relational contracts *at arm's length*, that is, that grant W an independent contractor's right to refuse to work for E. The reason is that because E and W have private information about their valuations, any transfers can only be based on observed actions. However, actions are already contractible, which leaves nothing for

a relational contract to improve upon.

To be more specific, there are two cases to distinguish, relational contracts in which "no work" is part of equilibrium play, and contracts where "no work" constitutes a breach of contract. If "no work" is a contractual option for W, then incentive compatibility implies that for any agreed transfer t^m in return for performing task m, W will choose m only if $c^m \leq t^m$. However, given the contractibility of actions, any such agreement is already feasible with spot trade. This would be different if work had a moral-hazard component, which could lead to an equilibrium as in Klein and Leffler (1981) in which the seller's (worker's) effort is motivated by repeat business.

If, however, "no work" is not an option for W within the contract, then the relationship is arguably some form of employment or equivalent long-term contract, because a worker who cannot decline to work for E is not truly an independent contractor. That said, this is a large class of contracts, leaving room for many ways for E and W to determine W's action. Within this class, I focus for the most part on a "traditional" employment contract that gives E the authority to choose W's action. In Section 6, I look at the opposite case in which W has the authority to choose.

4 One task

A model with only one task suffices to derive several main insights of the paper. In this and subsequent sections, I assume for the bargaining mechanisms (m^i, t^i) that with spot trade, E makes a take-it-or-leave-it offer to W, and with employment, W makes a take-it-or-leave-it offer to E. That is, the party that does *not* have control over W's actions has all bargaining power. Equivalently, since the party in control picks its best default option, an offer is made by the (opposite) party that is willing to *pay* for switching to a different action.[15] Aside from making the analysis tractable even with multiple tasks, this assumption biases the profit difference ΔV in favor of spot trade: with spot trade,

[15] Baker et al. (2013) make the same assumption in an extension their model that allows for costly ex-post negotiation. Similarly, rent-seeking models assume that "giving someone authority means that she will get lobbied" (Gibbons 2005) by those not in control, except that in those models, rent-seeking efforts are wasteful expenditures (e.g., on signal jamming), whereas here, one party directly offers a payment to another in return for a change in the decision.

no mechanism can give E a higher profit than a take-it-or-leave-it offer to W.[16] With employment, in turn, giving W all bargaining power minimizes E's profit relative to other mechanisms inasmuch as any adjustment rents earned by W can only partially be captured by E.[17] The results derived below therefore provide a lower bound to the relative profitability of employment for E.

4.1 Setup, spot trade, and employment without spot adjustments

In each period, W can either work for E ($m = 1$) or not work ($m = 0$). E's resulting benefit is mb with $b \in \{b_H, b_L\}$, where $b_H > b_L$ and $\Pr(b = b_H) = \beta$. W's cost is mc with $c \in \{c_L, c_H\}$, where $c_L < c_H$ and $\Pr(c = c_L) = \gamma$. Let $\Delta c = c_H - c_L$, $\Delta b = b_H - b_L$, $\overline{c} = E(c) = \gamma c_L + (1-\gamma)c_H$, and $\overline{b} = E(b) = \beta b_H + (1-\beta)b_L$. I assume $c_L < b_L < c_H < b_H$, which means that trade is efficient with positive probability and inefficient with positive probability. For this case, Bolton and Dewatripont (2005, Section 7.2.1) show that an efficient bargaining mechanism may not exist, as it never does with overlapping continuous types, per Myerson and Satterthwaite (1983). When E makes a take-it-or-leave-it offer to W, we obtain

Lemma 1. *Let* $\widehat{\gamma} = (b_H - c_H)/(b_H - c_L)$.

(a) If $\gamma \leq \widehat{\gamma}$*, E's optimal spot-trade prices are* $t(b_L) = c_L$ *and* $t(b_H) = c_H$*. Trade is efficient* ($S^0 = S^*$) *and E's profit is* $V^0 = S^* - \beta\gamma\Delta c$.

(b) If $\gamma > \widehat{\gamma}$*, E's optimal prices are* $t(b_L) = t(b_H) = c_L$*. Trade is inefficient* ($D^0 = \beta(1-\gamma)(b_H - c_H)$ *and* $S^0 = S^* - D^0$) *and E's profit is* $V^0 = S^0$.

The result is straightforward. E-type b_L can trade only with worker c_L, and thus optimally offers $t = c_L$. Type b_H can either trade with both worker types at $t = c_H$, or trade only with type c_L at $t_L = c_L$. If low-cost workers are more rare (low γ), then E opts for the higher price. In that case trade is efficient but W earns a rent Δc if $(b, c) = (b_H, c_L)$.

[16] In terms of Bajari and Tadelis's (2001) one-shot random-offers bargaining game in which the buyer (seller) makes an offer with probability λ $(1-\lambda)$, I assume $\lambda = 1$ for spot trade and $\lambda = 0$ for negotiations within employment. Assuming any fixed λ for both situations, instead, would shift the bargaining power to reduce the difference between E's profit from employment and from spot trade.

[17] With multiple tasks the effect is ambiguous, however, because W may be better informed than E once he has received E's order, which could make bargaining more efficient if W has the bargaining power (to E's benefit, by reducing D^E in (9)).

For larger γ, E opts for the low price. In that case, W earns no rent but the parties fail to trade if $(b, c) = (b_H, c_H)$.

Lemma 2. *Employment without spot adjustments leads to an expected deadweight loss of* $D^E = (1-\beta)(1-\gamma)(c_H - b_L)$*; thus* $S^E = S^* - D^E$*. E's profit is*

$$\widetilde{V}^E = S^* - D^E - (1-\delta)(c_H - c^E) \tag{11}$$

$$= S^* - (1-\beta)(1-\gamma)(c_H - b_L) - (1-\delta)\gamma\Delta c. \tag{12}$$

With employment, E's control over W's action simply means that W is relationally obligated to work for E in each period, which is efficient except in the state (b_L, c_H). Combining Lemmas 1 and 2, we obtain

Proposition 3. *Let* $\Delta S = S^E - S^0$*. Then in the binary model with* $c_L < b_L < c_H < b_H$ *and without spot adjustments within employment,*

$$\begin{aligned} &\textit{If } \gamma \leq \widehat{\gamma}\textit{:} \quad \Delta S = -D^E, \quad \Delta V = U^0 - D^E - (1-\delta)\gamma\Delta c, \\ &\textit{If } \gamma > \widehat{\gamma}\textit{:} \quad \Delta S = D^0 - D^E, \quad \Delta V = D^0 - D^E - (1-\delta)\gamma\Delta c, \end{aligned}$$

where $U^0 = \beta\gamma\Delta c$*,* $D^0 = \beta(1-\gamma)(b_H - c_H)$*, and* $D^E = (1-\beta)(1-\gamma)(c_H - b_L)$*. In particular,*

(1) ΔS *and* ΔV *are strictly increasing in* β*, and weakly increasing in* b_H *and* b_L*.*

(2) ΔS *and* ΔV *are strictly decreasing in* $\bar{c}$*, holding* Δc *constant (i.e.,* $\frac{d\Delta S}{dc_H} + \frac{d\Delta S}{dc_L} \leq 0, \frac{d\Delta V}{dc_H} + \frac{d\Delta V}{dc_L} \leq 0$*).*

Proof: See the Appendix. Proposition 3 states that, broadly speaking, both the difference in surplus and the difference in E's profit between employment and spot trade are increasing in the gains from trade between E and W. This result is driven by the different default outcomes with each governance mode: if for W to work is more efficient on average than not to work ($D^E < D^0$), it is efficient to move the default outcome to work, which is what employment accomplishes.

Because b and c represent *opportunity* benefits and costs, the surplus at stake equals the appropriable quasi-rent between E and W. Proposition 3 thus formalizes a central prediction of transaction-cost theory, that "larger AQRs make integration more likely" (Gibbons 2005). The bargaining framework used here makes precise and intuitive what

drives the result: Spot trade tends to be inefficient if work is mostly efficient but no work is the disagreement point. Employment improves on this not by suppressing "haggling" (see the next subsection) but by shifting the default to a more efficient outcome where further bargaining is less likely to be mutually beneficial.

As discussed in Section 3, if W is wealth-constrained, then ΔV is influenced by two additional effects, the obedience wage premium $(1-\delta)\gamma\Delta c$, which reduces the employment profit, and the rent earned by W as contractor, which reduces E's profit from spot trade.

Example 1. *Suppose* $(b_H, c_H, b_L, c_L) = (10, 8, 6, 4)$, $\beta = \frac{3}{4}$, *and* $\gamma = \frac{1}{4}$. *Then the first-best surplus is* $S^* = \frac{19}{8} = 2.375$.

(a) With spot trade, because $\gamma < \widehat{\gamma} = \frac{1}{3}$, *E offers* $t(b_H) = 8$ *and* $t(b_L) = 4$. *Trade is efficient, but E pays a rent to W in the state* (b_H, c_L) *with expected value* $U^0 = \frac{3}{4}$, *resulting in a profit of* $V^0 = S^* - U^0 = \frac{13}{8} = 1.625$. *This profit splits into an expected benefit* $E(b) = \frac{63}{8} = 7.875$ *and an expected cost* $E(t) = \frac{50}{8} = 6.25$.

(b) Employment is inefficient in the state (b_L, c_H), *resulting in expected deadweight loss* $D^E = \frac{3}{8}$ *and thus a surplus of* $S^E = 2$. *E's expected benefit is* $\overline{b} = 9$; *W's wage is, per (8),* $w = \overline{c} + (1-\delta)\gamma\Delta c = 8 - \delta$, *and E's profit therefore* $\overline{b} - w = 1 + \delta$.

(c) Applying (5), employment is feasible if $\delta \geq \frac{4}{5}$, *but is always less efficient than spot trade. Using Proposition 3,* $\Delta V = \frac{3}{4} - \frac{3}{8} - (1-\delta)\frac{1}{4}4$, *which is nonnegative for* $\delta \geq \frac{5}{8}$, *which is smaller than* $\frac{4}{5}$. *It follows that whenever employment is feasible, it is more profitable than spot trade.*

In the example, both E's benefit and her cost (the wage) are higher with employment than with spot trade. However, as $\delta \to 1$, the obedience premium shrinks and the wage converges to W's expected cost $\overline{c} = 7$, enabling E to capture all of the employment surplus in the limit. With spot trade, however, W always earns a rent of $\frac{3}{4}$, which is why, for large enough δ, employment can be more profitable than spot trade even if it is less efficient.

Last but not least, Proposition 3 also illustrates that when workers are wealth-constrained, the profitability of employment does not hinge on any information advantage on part of E, in contrast to a view that runs through most of the literature on employment.[18] To see this, consider the limit case $\beta = 1$, where the types reduce to $b_H > c_H > c_L$. In this case, only W knows the full state of the world, and yet employment may be profitable for the

[18] See, for example, Bolton and Rajan (2003), Wernerfelt (2004), Hart and Moore (2008), Marino et al. (2010), and Rantakari (2021), as well as Baron and Kreps (2013). However, the idea that a boss's information is the rationale for her authority goes back to Knight, Coase, Barnard, and Simon.

same reasons as discussed above: With spot trade, *any* incentive-compatible bargaining mechanism will lead to either a rent for W or a deadweight loss or both, i.e. $U^0 + D^0 > 0$ regardless of δ, whereas with employment, W's rent shrinks to zero as as $\delta \to 1$.

4.2 Employment with spot adjustments

Now suppose spot adjustments are feasible. Because (b_L, c_H), where work is inefficient, is the only state in which there are gains from trade, W with $c = c_H$ can (without knowledge of E's type) offer $t = b_L$ in order to switch to no work. For W to be able to finance this payment out of his current wage requires $w \geq b_L$, a condition that I investigate below. E-type b_H rejects the offer but b_L accepts, which restores efficiency:

Lemma 3. *If spot adjustments are feasible and $w \geq b_L$, then employment is efficient ($S^E = S^*$) and W's expected gain from spot adjustments is $U^E = (1-\beta)(1-\gamma)(c_H - b_L)$. E's profit is*

$$\widetilde{V}^E = S^* - (1-\delta)[\gamma\Delta c + (1-\beta)(1-\gamma)(c_H - b_L)]. \tag{13}$$

Combining Lemmas 1 and (13), we obtain

$$\begin{aligned} &\text{If } \gamma \leq \widehat{\gamma}\text{:} \quad \Delta V = U^0 - (1-\delta)(\gamma\Delta c + U^E) \\ &\text{If } \gamma > \widehat{\gamma}\text{:} \quad \Delta V = D^0 - (1-\delta)(\gamma\Delta c + U^E), \end{aligned} \tag{14}$$

which have the same properties as described in Proposition 3.

The fact that with spot adjustments, employment is first-best, whereas spot trade may not be, is partly an artifact of the small type space.[19] However, it also more generally reflects the fact that bargaining under incomplete information has a status quo bias. McKelvey and Page (2002) show this status quo bias formally for a generalization of Myerson and Satterthwaite's (1983) model; for a similar result, see Segal and Whinston (2016). It is a general feature of bargaining games and mechanisms: if no trade is the default, inefficiency takes the form of insufficient or delayed trade, and if trade is the

[19] Even with only three types for each party, employment may no longer be first-best. For instance, suppose $b = \overline{b} \pm \Delta$ and $c = \overline{c} \pm \Delta$, where $\overline{b} \geq \overline{c}$ and $\Delta > \overline{b} - \overline{c}$ ensure that benefits and costs are nested as in the binary model. If all types are equiprobable, then both with spot trade and employment, optimal prices are such that the party making an offer forgoes trade with its closest neighbor. In particular, with employment, W-type $c = \overline{c} + \Delta$ will only offer $t = \overline{b} - \Delta$ to switch to no trade, which type $\overline{b} - \Delta$ accepts but type $\overline{b}$ rejects, leading to a deadweight loss in the form of excessive trade. Details of this 3x3 model are available upon request.

default, inefficiency takes the form of excessive trade. It is then ex-ante efficient to assign the control right to the party that makes the most efficient decisions ex-post in the absence of an agreement, as Coase (1960) argued. And although authority does not by assumption supersede the use of prices, their use may be rare in equilibrium if E's orders are close to efficient already.

Applying this logic to the present model, since trade (work) is efficient in three of four states, and employment designates work as the default, it is easy for E and W to determine if work is inefficient, and negotiate away from work. By contrast, because spot trade designates no work as the default, which is inefficient in three of four states, it is relatively harder, due to incomplete information, for E and W to agree on efficient trade, and in equilibrium it may not always occur.[20]

Let us turn to W's financial constraint $t \leq w$, which here means $b_L \leq w$. If spot adjustments are feasible, then E's optimal wage according to (8) is

$$\begin{aligned} w &= c^E + (1-\delta)(c_H - c^E) - \delta U^E \\ &= \overline{c} + (1-\delta)\gamma\Delta c - \delta(1-\beta)(1-\gamma)(c_H - b_L) =: w_0(\delta). \end{aligned}$$

If $w < b_L$, however, and therefore $U^E = 0$, (8) leads to the wage

$$w = \overline{c} + (1-\delta)\gamma\Delta c =: w_1(\delta) > w_0(\delta),$$

which is strictly higher than $w_0(\delta)$, and E's profit therefore strictly lower. Because of this discontinuity it may be optimal to hold w at b_L over some range of δ, to enable W to finance spot adjustments away from inefficient work. For the next result, recall that $b_L > c_L$, and that employment is feasible if $\delta \geq \delta_0$.

[20] Incidentally, the ordering of types $b_L < c_L < c_H < b_H$ is realistic too: a service may always be costly to the seller, but sometimes valued only little by the buyer. Bolton and Dewatripont (2004, page 244, Footnote 3) point out that in this case, an efficient mechanism is easy to determine: agree on a price $p \in [c_H, b_H]$ and let E decide if she wants to buy at this price. In my model, however, this case is barely different from the main case considered as far as E's profit is concerned: spot trade looks exactly the same, resulting in either a rent for W or a deadweight loss. With employment, requiring work when $b = b_L$ is now always inefficient, but again W can pay E to switch to no work.

The case $c_L < b_L < b_H < c_H$ is different, however. Only trade with c_L is efficient, and with spot trade, both types of E can reach the first-best by offering $t = c_L$. Authority then has no advantage. What the other two cases have in common, therefore, is that irrespective of the bargaining mechanism used, E is unable to capture the full first-best surplus through spot trade. In particular, any efficient mechanism necessarily leaves a rent to W. This is what creates an opening for employment to lead to higher profits when W lacks wealth.

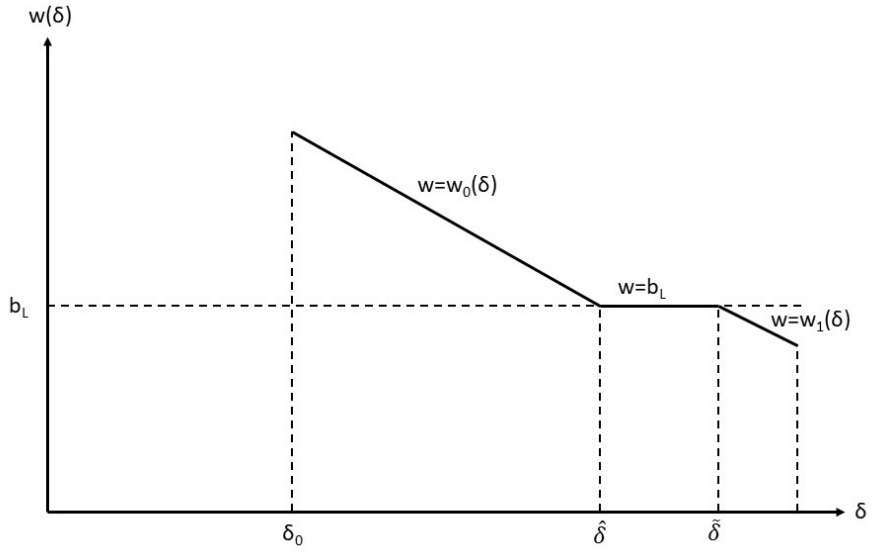


Figure 2: Optimal wage as function of δ for case (c) of Proposition 4

Proposition 4. *(a) If $b_L \in (c_L, \overline{c} - U^E]$, then E's profit-maximizing wage is $w = w_0(\delta)$ for all $\delta \in (\delta_0, 1]$, and spot adjustments are feasible for all δ.*

(b) If $b_L \in (\overline{c} - \delta U^E, \overline{c}]$, then there exists $\widehat{\delta} \in (\delta_0, 1]$ such that the optimal wage is $w = w_0(\delta)$ for $\delta \in (\delta_0, \widehat{\delta}]$, and $w = b_L$ for $\delta \in (\widehat{\delta}, 1]$.

(c) If $b_L > \overline{c}$, then there exists $\widetilde{\delta} \in (\widehat{\delta}, 1]$ such that the optimal wage is $w = w_0(\delta)$ for $\delta \in (\delta_0, \widehat{\delta}]$, $w = b_L$ for $\delta \in (\widehat{\delta}, \widetilde{\delta}]$, and $w = w_1(\delta)$ for $\delta \in (\widetilde{\delta}, 1]$.

Case (c), for low values of b_L, is shown in Figure 2. W's obedience constraint requires the wage to be at least $w_0(\delta)$, which is optimal as long as $w_0(\delta) \geq b_L$. If $w_0(\delta) < b_L$, then it is optimal to set $w = b_L$ as long as $w_1(\delta)$, the required wage without adjustments, exceeds b_L. Only if $w_1(\delta) < b_L$ is it optimal to forgo spot adjustments and set $w = w_1(\delta)$.

In the example of Section 4.1, however, the simplest case (a) of Proposition 4 applies, and therefore Lemma 3 holds too:

Example 2. *Consider Example 1 but now suppose that spot adjustments within employment are feasible. Employment is then efficient, $S^E = S^* = \frac{19}{8} = 2.375$, and the deadweight loss from Example 1 is captured by W as gain from spot adjustments, $U^E = \frac{3}{4}$. W's wage is $w_0(\delta) = 7 + (1 - \delta) + \delta\frac{3}{8} = 8 - \frac{11}{8}\delta$, therefore $w_0(1) > b_L$ and case (a) of Proposition 4 applies.*

Given the greater surplus S^E, employment is feasible for a larger range of δ than before, namely $\delta_0 \approx 0.77$ instead of 0.8. Using (14), we have $\Delta V = \frac{3}{4} - (1 - \delta)(1 + \frac{3}{8})$, which

is nonnegative for $\delta \geq \frac{5}{11}$. It follows that whenever employment is feasible, it is strictly more profitable than spot trade, even though both spot trade and employment are efficient.

Allowing for spot adjustments is both realistic and methodologically appealing, as doing so places spot trade and employment on the same timeline and highlights a key argument of the paper: the advantage of employment is not to supplant bargaining with authority, but to shift the default outcome to a more efficient action than the default market outcome, no trade.

In addition, the fact that the drivers of ΔV are similar without and with spot adjustments shows that the feasibility of spot adjustments is not critical to this paper's most important results. To avoid uninteresting case distinctions, I assume in subsequent sections that spot adjustments *are* feasible and are affordable for W. This amounts to assuming, as in Example 2, that b_L is small enough relative to W's costs for any spot transfers to E to be fundable out of W's wage.

5 Multiple tasks

The idea that employment enables a boss to direct a worker between *different* actions goes back to Coase (1937) and runs through the entire literature on employment. To examine this idea, this section expands the model of Section 4 to two tasks.

5.1 Setup and equilibria

Suppose now that there are two mutually exclusive tasks X and Y. With spot trade, E offers prices (t^X, t^Y). W can accept at most one of them and perform the corresponding task, or reject both prices and not work for E. With employment, E orders W to perform the task that leads to the highest benefit for her. W can offer prices to switch to either the other task, or to no work, which E accepts or rejects.

E's and W's valuations for each task are the same as in Section 4: $b^i \in \{b_L, b_H\}$ and $c^i \in \{c_L, c_H\}$ for $i \in \{X, Y\}$, with $c_L < b_L < c_H < b_H$. Each party's valuations may be correlated across tasks. Given symmetry as assumed in Section 3, let $p_H = \Pr(b_H, b_H)$, $p_L = \Pr(b_L, b_L)$, and $p_M = \Pr(b_H, b_L) = \Pr(b_L, b_H)$; and $q_H = \Pr(c_H, c_H) = q_H$, $q_L = \Pr(c_L, c_L)$, and $q_M = \Pr(c_H, c_L) = \Pr(c_L, c_H)$. For comparison with the single-task model,

E's type:	(b_H, b_H)	(b_H, b_L)	(b_L, b_L)	U^0	D^0
If $\Delta b \geq \Delta c$:	E's optimal price (t_H, t_L):				
$\gamma \leq \gamma_0$:	(c_H, c_L)	(c_H, c_L)	(c_L, c_L)	$(1-p_L)\gamma\Delta c$	0
$\gamma \in (\gamma_0, \gamma_1]$:	(c_L, c_L)	(c_H, c_L)	(c_L, c_L)	$2p_M\gamma\Delta c$	$p_H q_H(b_H - c_H)$
$\gamma > \gamma_1$:	(c_L, c_L)	(c_L, c_L)	(c_L, c_L)	0	$(1-p_L)q_H(b_H - c_H) + 2p_M q_M(\Delta b - \Delta c)$
If $\Delta b < \Delta c$:	E's optimal price (t_H, t_L):				
$\gamma \leq \gamma_0$:	(c_H, c_L)	(c_H, c_L)	(c_L, c_L)	$(1-p_L)\gamma\Delta c$	0
$\gamma > \gamma_0$:	(c_L, c_L)	(c_L, c_L)	(c_L, c_L)	0	$(1-p_L)q_H(b_H - c_H)$

Table 1: Prices, W's payoff, and deadweight loss with spot trade over two tasks

assume that the marginal distributions over b^i and c^i are the same as as in Section 4, that is, $p_H + p_M = \beta$, $p_L + p_M = 1 - \beta$, $q_L + q_M = \gamma$, and $q_H + q_M = 1 - \gamma$.

If $p_M = 0$, b^X and b^Y are perfectly correlated, whereas if $p_M = \beta(1-\beta)$, they are uncorrelated. If $p_M > \beta(1-\beta)$, then b^X and b^Y are negatively correlated. For $\beta \leq 1/2$, the maximum negative correlation is given by $p_M = \beta$, in which case $p_H = 0$ and $p_L = 1-2\beta$, whereas if $\beta \geq 1/2$, the maximum negative correlation is $p_M = 1-\beta$, in which case $p_L = 0$ and $p_H = 2\beta - 1$. Similarly, $q_M \in [0, \max\{\gamma, 1-\gamma\}]$ measures (inversely) the correlation between c^X and c^Y.

As discussed in Section 3, both E's orders and bargaining under spot trade and employment select for Pareto-efficient outcomes. Thus, if E is indifferent between tasks X and Y but W is not, then E will order the lowest-cost task. With spot trade, if for instance $b = (b_H, b_L)$ and $t = c = (c_H, c_L)$, W is indifferent between the tasks but E is not if $\Delta b \neq \Delta c$, and the bargaining mechanism selects the more efficient option.

Lemma 4. *For $\widehat{\gamma}$ as defined in Lemma 1, let $\gamma_0 = (1-q_M)\widehat{\gamma}$ and $\gamma_1 = \widehat{\gamma} - \frac{b_L - c_L}{b_H - c_L} q_M$, where $\gamma_0 \leq \gamma_1 \leq \widehat{\gamma}$. Then for spot trade, E's optimal prices, W's expected payoff U^0, and the expected deadweight loss D^0 are shown in Table 1. In particular, for any distribution of W's costs (and the corresponding value of γ), E's average price per task is lower than with only one task.*

Proof: see the Appendix. As in the general model, E's spot-trade profit is $V^0 = S^* - U^0 - D^0$, where S^* is larger than with only one task because there are two tasks presenting opportunities for trade. E's prices are unambiguously lower than with one task because E can gamble that W will accept a low price for one *or* the other task.

Prices, W's payoff, and the deadweight loss vary with γ in the same way as with one

task, except that now there are up to three cases to distinguish. If γ is small ($\gamma \leq \gamma_0$), E prices conservatively, and there is no deadweight loss but W earns a rent. If γ is large ($\gamma > \gamma_1$), E prices most aggressively, W earns no rent but there is a deadweight loss. Part of the deadweight loss is forgone trade as before.

A new source of potential inefficiency, reflected in the term $2p_M q_M(\Delta b - \Delta c)$, arises if $\Delta b \geq \Delta c$: spot trade will lead to the *wrong* task being ordered if, due to E's low prices, W chooses a task with $(b^i, c^i) = (b_L, c_L)$ even though the other task (with $(b^j, c^j) = (b_H, c_H)$) is more efficient. If $\Delta b < \Delta c$, then W's choices are always efficient.

Lemma 5. *With two tasks, employment is efficient ($D^E = 0$), and total surplus is $S^E = S^* = b^E - c^E + U^E$, where*

$$\begin{aligned} b^E &= (1-p_L)b_H + p_L b_L, \\ c^E &= \overline{c} - (p_L + p_H)q_M \Delta c, \text{ and} \\ U^E &= \begin{cases} p_L q_H (c_H - b_L) & \text{if } \Delta b \geq \Delta c, \\ p_L q_H (c_H - b_L) + 2p_M q_M (\Delta c - \Delta b) & \text{if } \Delta b < \Delta c. \end{cases} \end{aligned}$$

E's profit is $\widetilde{V}^E = S^* - (1-\delta)(c_H - c^E + U^E)$.

Proof: see the Appendix. With employment, E's expected benefit before adjustments, b^E, is not simply $\overline{b}$ like with one task, but instead $b^E = E[\max\{b^X, b^Y\}] > \overline{b}$. And because in case of indifference, E orders the lower-cost task, W's expected cost c^E is not $\overline{c}$ but smaller. Like in Section 4.2, employment is efficient because, out of 16 states of the world, E's order is inefficient in either one or three states. One of them is the state $(b, c) = (b_L, b_L, c_H, c_H)$. Here, although W does not know E's valuations, type $c = (c_H, c_H)$ can offer $t = b_L$ to E in order to switch to no work, which E accepts if and only if $b = (b_L, b_L)$, thus restoring full efficiency.

The ranking of Δb and Δc matters for what happens when $b = (b_H, b_L)$ and $c = (c_H, c_L)$ or vice versa. If $\Delta b \geq \Delta c$, then E's most profitable task, $\arg\max_{X,Y}\{b^X, b^Y\}$, is also the efficient task. If $\Delta b < \Delta c$, however, E will order the wrong task. Inefficiency is still avoided in this case because W can negotiate a switch to the other task by offering $t = \Delta b$ to E (and gain $\Delta c - \Delta b > 0$), if (1) his costs are different and (2) E selects the task with the higher cost. Due to the tie-breaker rule that E already selects the most cost-efficient task if she is indifferent, these two conditions pin down E's type, out of four possible ones.

Thereby, the first-best is attained and W pockets the full gain from adjustment.

The latter result illustrates a general advantage that employment has over spot trade. Not only does the exercise of authority lead to unilateral adaptation (according to E's valuation) as opposed to no adaptation. In addition, with multiple tasks, E's order conveys information about her type, which makes it easier for W to target his adjustment price offers, reach an efficient outcome, and capture the gain. By contrast, with spot trade, because the default is no work, the parties have not learned anything when they negotiate spot prices, which makes bargaining relatively less efficient.

5.2 Do multiple tasks favor authority?

Like with one task, employment is both more efficient and more profitable the greater the surplus (the appropriable quasi-rent) at stake:

Proposition 5. *For all cases covered by Lemmas 4 and 5, and both with and without spot adjustments within employment,*

(a) ΔS *and* ΔV *are strictly increasing in* $\overline{b}$*, holding* Δb *constant (i.e.,* $\frac{d\Delta S}{db_H} + \frac{d\Delta S}{db_L} \geq 0$, $\frac{d\Delta V}{db_H} + \frac{d\Delta V}{db_L} \geq 0$*).*

(b) ΔS *and* ΔV *are strictly decreasing in* $\overline{c}$*, holding* Δc *constant (i.e.,* $\frac{d\Delta W}{dc_H} + \frac{d\Delta W}{dc_L} \leq 0$, $\frac{d\Delta V}{dc_H} + \frac{d\Delta V}{dc_L} \leq 0$*).*

The similarity with Proposition 3 reflects two senses in which the nature of employment does not hinge on a multiplicity of tasks. First, the drivers of ΔV are the same three as with one task: the relative efficiency of employment, which depends on the surplus at stake, the obedience wage premium, and E's ability to better minimize W's rent with employment.

Second, having two tasks instead of one does not systematically increase either ΔS or ΔV, in spite of an increase in S^*: With spot trade, both D^0 and U^0 can be larger or smaller with two tasks than with one. With employment, W's adjustment gain U^E may be smaller or larger than with one task depending on $\Delta b \lesseqgtr \Delta c$. W's default cost c^E, finally, is lower than the ex-ante average $\overline{c}$, which raises the obedience wage premium is $(1-\delta)(c_H - c^E)$. Accordingly, it is easy to construct examples where the presence of a second task can increase or decrease ΔV; see Section 5.3.

Coase (1937) pointed out that "if a workman moves from department Y to department X, he does not go because of a change in relative prices, but because he is ordered to do

so." Accurate though Coase's observation may be empirically, it does not establish that authority is efficient—the same workman as contractor would likely move from Y to X if faced with prices $t^X > t^Y$. Thus, Alchian and Demsetz's (1972) assertion that "telling an employee to type this letter rather than to file that document is like my telling a grocer to sell me this brand of tuna rather than that brand of bread" has a clear formal expression in my model: If $b = (b_H, b_L)$, then with employment, E orders W to perform task X, whereas in the market, she can direct W by offering prices $t = (c_H, c_L)$—unless she chooses $t = (c_L, c_L)$ to gamble on W having a low cost. The latter behavior might be labeled opportunistic haggling (although by the buyer not the seller!) or costly discovery of "what the relevant prices are" (Coase). However, it already appears in the one-task model of Section 4.[21]

5.3 Correlation of costs and benefits, and job design

The profitability of employment depends on how E's and W's valuations are correlated across tasks, which may be the real reason why employees often have many different tasks:

Proposition 6. *(a) For p_H, p_L, $p_M > 0$, consider an increase in the correlation of b^X and b^Y that leaves their marginal distibutions unchanged, in the form of a simultaneous increase in both p_H, p_L by ε_b and a decrease in p_M by ε_b. Then ΔV is decreasing in ε_b.*

(b) For q_H, q_L, $q_M > 0$, consider an increase in the correlation of c^X and c^Y in the form of a simultaneous increase in both q_H, q_L by ε_c and a decrease in q_M by ε_c. Then ΔV is increasing in ε_c.

Proof: see the Appendix. Per Proposition 6, a lower correlation of benefits unambigously increases the profit difference $\Delta V = \widetilde{V}^E - V^0$: when E strictly prefers one task over the other, spot trade is more likely to lead to a rent for W or a deadweight loss. Meanwhile, with employment, E's default orders are more likely to be efficient, and the obedience wage premium is smaller because W's expected cost is closer to the ex-ante

[21] In the literature, reasons for why the ability to direct a worker to different tasks is essential to authority take different forms. Coase (1937) appealed to direct costs of bargaining, arguing that it is less costly to negotiate one wage than prices for every possible task. Simon (1951) compared employment not with ex-post trade, but with an ex-ante agreement on a specific task. For employment to be different from such an agreement requires at least two tasks. The papers by Van den Steen (2010) and Rantakari (2021), finally, focus on coordination problems, not trade, which by definition require at least two distinct actions.

value $\overline{c}$. A higher correlation of W's costs has exactly the same three effects.[22]

Proposition 6 captures most closely the intuition expressed by Coase, Simon and others that employment is more likely to be optimal the less the worker cares which task he performs, and the more the boss (E) cares.[23] A negative correlation between b^X and b^Y means that E has high demand for X in some periods and for Y in others, or (more loosely interpreted) can shift these actions across time to achieve a negative correlation. As a result, E has a high valuation for W to perform one or the other action in most periods, as reflected in the fact that $d(\max\{b^X, b^Y\})/dp_M > 0$.

A positive correlation between c^X and c^Y, in turn, means that W cares relatively little which of the two tasks he performs. The twist here is that W still cares about whether he works at all. The distinction between the extensive margin (whether to work) and the intensive margin (which task to work on) has not been made in the literature but is critical for understanding employment and alternative market arrangements. The Uber drivers studied by Chen et al. (2019), for instance, are found to have greatly fluctuating opportunity costs of working. When they do work, however, their preferences over destinations presumably vary little over time. Put differently, when changes in the worker's costs are driven mainly by outside opportunities, the costs are likely to be highly correlated across tasks.

Proposition 6 has consequences for job design, as it implies that bundling tasks with the right correlation structure into one job may *make* employment more profitable:

Example 3. *(a) Consider first a job with a single task with* $(b_H, c_H, b_L, c_L) = (10, 8, 6, 4)$*, and* $\beta = \gamma = \frac{1}{2}$*. Then* $\gamma > \widehat{\gamma}$*, and spot trade leads to* $U^0 = 0$ *and* $D^0 = \frac{1}{2}$ *according to Lemma 1. From Lemma 3,* $U^E = \frac{1}{2}$ *and therefore* $S^E = \overline{b} - \overline{c} + U^E = \frac{5}{2}$*. Employment is feasible if* $\delta \geq \delta_0 = \frac{16}{21} \approx 0.76$*, and is preferred over spot trade if* $\Delta V = \frac{1}{2} - (1-\delta)(\frac{1}{2}4 + \frac{1}{2}) \geq 0$*, which is the case if* $\delta \geq \frac{4}{5} = 0.8$*.*

(b) Now suppose there are two identical tasks X and Y with the same costs, benefits, and marginal distributions as above, and consider two different correlation structures. If the costs and benefits are uncorrelated, then with spot trade (since $\gamma > \widehat{\gamma} > \widehat{\gamma}_1$*),* $U^0 = 0$

[22] Proposition 6 does not unambiguously carry over to ΔS, which due to the efficiency of employment is driven only by the spot trade deadweight loss D^0: for $\gamma \in [\gamma_0, \gamma_1)$, a lower correlation of b^X and b^Y increases ΔV per the proposition, but decreases ΔS because spot trade becomes more efficient.

[23] Coase (1937: 391) writes "It may well be a matter of indifference to the person supplying the service or commodity which of several courses of action is taken, but not to the purchaser of that service or commodity." And Simon (1951: 295): "W will be willing to enter an employment contract with B obly if it does not matter to him 'very much' which x ... B will choose".

and $D^0 = \frac{3}{8}$. *With employment,* $b^E = 9 > \overline{b}$, $c^E = \frac{11}{2} < \overline{c}$, *and* $U^E = \frac{1}{8}$. *It follows that* $S^E = S^* = \frac{29}{8} = 3.625$, *and employment is feasible if* $\delta \geq \frac{64}{93} \approx 0.69 < 0.76$. *Employment is most profitable if* $\Delta V = \frac{3}{8} - (1-\delta)(8 - \frac{11}{2} + \frac{1}{8}) \geq 0$, *which is the case if* $\delta \geq \frac{6}{7} \approx 0.86 > 0.8$. *Thus, comparing with (a), adding a second identical task decreases the range of* δ *for which employment is more profitable than spot trade.*

(c) Suppose instead that the benefits are maximally negatively correlated with $p_M = \frac{1}{2}$ *and* $p_H = p_L = 0$, *and the costs are perfectly correlated, thus* $q_L = q_H = \frac{1}{2}$ *and* $q_M = 0$. *Then* $U^0 = 0$, $D^0 = 1$, $b^E = 10$, $c^E = 6$, *and* $U^E = 0$. *Then* $S^E = 4$, *and employment is feasible if* $\delta \geq \frac{2}{3}$. *Employment is most profitable if* $\Delta V = 1 - (1-\delta)(8-6) \geq 0$, *which is the case if* $\delta \geq \frac{1}{2}$. *Therefore, if employment is feasible at all, it is more profitable than spot trade, for a larger range of* δ *than with one task (*$\delta_0 = 0.66$ *vs.* 0.8*).*

(d) Could job (c) be split into two jobs, for an X and a Y specialist who perform their task when needed but who are not required to work ($c = 0$*) if not? Yes, but only at a strictly higher total wage cost and thus lower profit: for the bundled job in (c), the required wage is* $c^E + (1-\delta)(c_H - c^E) = 8 - 2\delta$. *If the job is split and each worker is idle if the other's task is demanded, each needs to be paid* $c^E/2 + (1-\delta)(c_H - c^E/2) = 8 - 5\delta$, *and* $2(8-5\delta)$ *exceeds* $8 - 2\delta$ *for any* $\delta < 1$.

Part (c) of Example 3 shows that while a firm may not have enough demand to hire different specialists to "type this letter" or to "file that document," it is much more likely to have demand for one worker to do one or the other at different times, especially if the worker cares little which task is requested. Moreover, part (d) shows that there are scale economies to paying for availability: even if specialists could stay at home (so to speak) if not needed and receive an accordingly lower wage, ensuring their availability by paying the requisite obedience premium leads to strictly higher wage costs than one bundled job.

These results suggest a causal link between employment and task multiplicity that differs from the one emphasized in the literature. Coase and many others argued that decision-making by authority is efficient when workers perform different tasks that cannot be specified ex-ante. This argument implicitly takes job design as given. Proposition 6 and Example 3, by contrast, suggest that *giving* a worker multiple tasks *makes* employment more profitable relative to spot trade because firms will pay a regular wage only to people who always have something to do. This argument provides a new economic rationale for how to bundle tasks into jobs in addition to familiar ones such as multitasking and

learning, see Lazear and Oyer (2013, Section 6.1).

A case in point is the job of a secretary, a favorite example of economists perhaps because it "strikes close to home" (Fama 1991, see also Alchian and Demsetz 1972, Simon 1991, Wernerfelt 1997) even though it is not the most typical job. Wernerfelt (1997), for instance, argues that it is efficient for a secretary to be employed because it would be too costly to negotiate spot prices for the many tasks a secretary performs. The results above suggest a different perspective: The position of a secretary only exists because managers, professors, and others have demand for services that individually would be too costly too outsource, due to low individual demand, but can (in terms of human-capital requirements) be performed by the same person who is paid to be available for whenever demand for a secretarial service arises.

6 Worker authority

Would E want to employ W but delegate the choice of tasks to him? Simon (1951) briefly mentions this possibility; in Marino et al. (2010), the link between obedience and delegation is the central question. In my model, employment with delegation is distinct from spot trade because the employee with authority is still required to provide a service whereas the contractor is not. This distinction of employee and contractor differs from Hart and Moore's (2008), who label a contract with buyer authority "employment" and a contract with seller authority "independent contracting", with neither contract entailing any obligation, as trade is assumed to be voluntary ex post.[24]

In spite of this difference, both Hart and Moore's results and Lemma 5 suggest that the optimality of E- vs. W-authority may hinge on whether $\Delta b \lesseqgtr \Delta c$. Indeed (as stated in the next result), *without* spot adjustments, E-authority is most efficient if $\Delta b \geq \Delta c$, and W-authority if $\Delta c > \Delta b$. Both entail inefficient trade in the state $(b, c) = (b_L, b_L, c_H, c_H)$ because W is required to provide his service either way. However, when $(b^X, c^X) = (b_H, c_H)$ and $(b^Y, c^Y) = (b_L, c_L)$ or vice versa, E's order is efficient if $\Delta b \geq \Delta c$ and W's own choice is efficient if $\Delta c < \Delta b$. Nevertheless, as shown in the next result, it tends to be more

[24]Hart and Moore consider the example of hiring a musician for a gathering and determining who should pick the music, the client or the musician. But for a one-time engagement, the musician would intuitively and legally be a contractor irrespective of who chooses. And if the client were to hire the musician as employee, it would likely be for reasons such as the frequency of demand, again irrespective of who chooses the music.

profitable for E to retain authority for herself.

In both this section and the next, either E or W may be willing to pay for spot adjustments within employment. Generalizing the bargaining assumptions made in Section 4, I assume that either side can offer a payment $t \geq 0$ to the other in return for a switch to a different action (i.e., this rules out offering a switch in return for *receiving* money from the other party). If only one side makes an offer, the other accepts or rejects. If E and W make simultaneous offers and they are compatible (suggesting the same switch and offering a nonnegative transfer to the other side), then an agreement is reached without payment, as the switch constitutes a Pareto improvement.

Proposition 7. *(a) If within employment, no spot adjustments are feasible, then W-authority is most efficient if and only if $\Delta b < \Delta c$. W-authority is most profitable if and only if $\Delta b < \Delta c$ and $\delta > \Delta b / \Delta c$.*

(b) With spot adjustments and nonnegative take-it-or-leave it offers by either side, both E- and W-authority are efficient, but E-authority is always more profitable.

Proof: see the Appendix. If $\Delta b < \Delta c$, then delegating authority to W lowers the cost of W's default actions, c^E, by more than it lowers E's benefit b^E, which is efficient as far as default actions are concerned. However, E's profit is $\widetilde{V}^E = b^E - w + V^E$ per (3), and especially when spot adjustments are feasible, the reduction in W's wage is much smaller than the decrease in c^E, as can be seen in (8): First, the reduction in direct costs c^E is offset by a tighter obedience constraint (4) and hence a larger obedience wage premium $(1 - \delta)(c_H - c^E)$. This explains why in the absence of spot adjustments (part a), delegation is optimal only if δ is sufficiently large (or equivalently, the worker sufficiently better informed, $\Delta c > \Delta b / \delta$). Second, with spot adjustments (part b), E's ability to partially recoup W's adjustment gains through a lower wage means that the greater efficiency of W's default actions only partially benefits E because it reduces W's adjustment gains. As the proof shows, the direct loss in E's benefit from delegating authority to W unambigously exceeds the saving on wage costs.

Proposition 7 contrasts with Marino et al.'s (2010) conclusion that delegation relaxes the worker's obedience constraint, and may thus be optimally chosen by the employer even though centralization would lead to better decisions. Here it is the other way around: delegation can only ever be optimal if the worker makes better decisions, but it may not be profitable because the obedience constraint is *tighter*. It is tighter because (i) W's highest

possible cost is still c_H even if W has authority, and (ii) W's lower costs are reflected in a lower wage, or put differently, W's utility gain from being in control is neutralized by a negative compensating differential. Effect (i) may be specific to this model's symmetry assumptions, but effect (ii) is likely to be present whenever workers compete for a job.

The possibility of worker authority suggests a more general definition of employment than Simon's (1951): *Employment is a relational contract that obligates a worker to work for a firm, in return for a wage. Whether the firm or the worker has control over the worker's actions is endogenous.* Proposition 7 may help explain why, historically, firm authority has been the norm and worker authority the exception.

Finally, Proposition 7 complements the implication of Proposition 3 that employment can be optimal even if E is not better informed than W about how best to adjust actions to the state of nature, contrary to most of the literature since Knight. Proposition 7 shows that E-authority can be optimal even if W-authority is an option. Moreover, even if W-authority is optimal, employment with delegation may dominate spot trade.

This result partially helps resolve a puzzle posed by Freeland and Zuckerman (2018): Why are hierarchies and employment still so dominant in the 21st century, even though in the knowledge economy, "managers cannot possibly hold sufficient knowledge to direct the firm efficiently because of the importance of adapting to changing local conditions"? Freeland and Zuckerman propose their own answer, but one that follows from my analysis is that managers' knowledge advantage is not that important a condition for employment to begin with.

7 Limits of authority?

No work since Simon (1951) has examined the optimal scope of a boss's formal authority.[25] Since in my model, employment is already well-defined with only one task (say X), we can ask: when should a second task Y be included in E's authority?

[25] Both Marino et al. (2010) and Rantakari (2021) study the limits of authority but from different angles. In Marino et al. (2010), obedience constraints may force a principal to delegate decisions to the agent; see Section 6. In Rantakari (2021), the threat of disobedience endogenously limits the superior's ability to tell the subordinate what to do, but there is no "area of acceptance" as in Simon (1951).

7.1 A task that is always more costly

Suppose that initially, $A = \{X\}$; i.e. employment requires W to perform task X but not Y. In contrast to the setup of Section 5.1, suppose that task Y is always more costly to W than X, but is sometimes efficient to perform, such as overtime work. Suppose that for task X, W's cost is c_L with probability γ and c_H with probability $1-\gamma$, whereas E's benefit is always b. For task Y, E's benefit is higher than b by either b_L (with probability $1-\beta$) or b_H (with probability β), whereas W's cost is higher than c_L or c_H by a constant amount c. The possible valuations are thus $b^Y \in \{b+b_L, b+b_H\}$ and $c^Y \in \{c_L+c, c_H+c\}$.

The costs c^X and c^Y are independent, and for simplicity I assume that each task is more efficient than no work; i.e. $b > c_H$ for X and $b + b_L > c_H + c$ for Y. However, it depends on the state of the world which task is most efficient. As shown in the next result, the assumption

$$b_L + \Delta c < c < b_H - \Delta c \tag{15}$$

ensures that task Y is efficient if and only if E's benefit is $b^Y = b + b_H$. I will compare three possible contracts:

1. A contract with $A = \{X, Y\}$ and a constant wage w.
2. A contract with $A = \{X\}$ but the possibility to renegotiate to task Y.
3. A contract with $A = \{X, Y\}$ but with a wage differentiated by task: $w = \{w^X, w^Y\}$.

Proposition 8. *(a) If (i) spot adjustments are feasible with nonnegative take-it-or-leave it offers by either side, (ii) S's costs are nonnegatively correlated across tasks, and (iii)*

$$4\Delta c \geq b_H - c, \tag{16}$$

then a contract with $A = \{X\}$ is more profitable than a contract with $A = \{X, Y\}$ and a constant wage w if

$$\beta q_M \Delta c < (1-\beta)(c - b_L). \tag{17}$$

(2) A contract with $A = \{X, Y\}$ and $w^Y = w^X + c$ is more profitable than either a contract with $A = \{X\}$ or a contract with $A = \{X, Y\}$ and a constant wage.

Proof: see the Appendix. Part (a) of Proposition 8 shows that excluding task Y from E's authority may be best if the wage cannot be tailored to the task. Per condition (17)

this is the case, quite intuitively, if β is sufficiently small; i.e. if task Y is rarely efficient. Condition (16) pins down the optimal price $t = c$ that E offers to W to switch to Y when $A = \{X\}$, and is imposed mainly to avoid uninteresting case distinctions.[26]

However, per part (b), both options are unambiguously dominated by a contract that includes both tasks but differentiates the wages. Remunerating task Y at a wage that is higher by c accomplishes two things: first, it induces E to order Y only if it is efficient. Second, it leads to the least restrictive obedience constraints for W, as a high cost is matched with a high wage.

Thus, one reason to limit E's authority is to exclude high-cost tasks if they end up being ordered too often when they are inefficient. Better still, however, may be to tie wages to actions, as is common practice with overtime work.[27]

7.2 Same support, different distribution

Suppose now that for task Y, the supports are the same as for X, $\{b_L, b_H\}$ and $\{c_L, c_H\}$, but the distributions different. I'll allow for arbitrary distributions $\Pr(b_i, b_j) = p_{ij} > 0$ and $\Pr(c_i, c_j) = q_{ij} > 0$ for $i, j \in \{H, L\}$, and $\Sigma\ p_{ij} = \Sigma\ q_{ij} = 1$. Now, under weak conditions, it is always optimal to include Y in the job:

Proposition 9. *If $\Delta b/\Delta c \in [p_{HL}, 1/q_{LH}]$, then employment with $A = \{X, Y\}$ and a uniform wage is always more profitable for E than a contract with $A = \{X\}$ and with possible spot adjustments to Y.*

Proof: see the Appendix.[28] The stated parameter constraint allows for $\Delta b \lesseqgtr \Delta c$ but requires the ratio to be close enough to 1. Like in the previous result, it pins down the price that either E or W offers to the other to switch from X to Y if $A = \{X\}$.

Proposition 9 states, strikingly, that it is *never* optimal for E to exclude a task whose support of valuations is the same as those of included tasks, no matter how inefficient

[26] Parameters satisfying all assumptions made in this section exist, for example, $b = 5$, $b_L = 1$, $b_H = 5$, $c_H = 3$, $c_L = 2$, and $c = 2.5$.

[27] Although not part of his formal analysis, Simon (1951) recognized this possibility: "W will be willing to enter an employment contract with B only if it does not matter to him "very much" which x ... B will choose or if W is compensated in some way for the possibility that B will choose an x that is not desired by W (i.e., that B will ask W to perform an unpleasant task)." (page 295).

[28] A uniform wage is not guaranteed to be optimal because allowing for asymmetric distributions violates the assumptions of Proposition 1. That said, differentiated wages $w^X \neq w^Y$ are unlikely to be optimal as they would both unnecessarily relax one of W's obedience constraints and likely distort E's optimal order. In any case, however, the restriction strengthens the result.

task Y may be on average, as long as it is sometimes efficient. That is, the assumption $p_{ij}, q_{ij} > 0$ ensures that Y is efficient in at least one of the 16 states of the world, but $\Pr(b^Y - c^Y > b^X - c^X)$ can be arbitrarily small, and the ex-ante surplus of certain trade $\overline{b}^Y - \overline{c}^Y$ can be negative.

The result is, like Lemma 5 for $\Delta b < \Delta c$, driven by the fact that if $A = \{X, Y\}$, E's order conveys information about her type that enables W to target adjustment offers to E very precisely. For instance, even in the case $\Delta b < \Delta c$ where E sometimes orders the wrong task, W can infer E's type and renegotiate to the efficient task.

By contrast, when $A = \{X\}$, E always orders X and renegotiating to Y may fail because W does not learn anything about E's type. For instance, if $\Delta b < \Delta c$, $b = (b_H, b_L)$ and $c = (c_H, c_L)$, then E orders X but Y is more efficient. W could offer $t = \Delta b$ for E to switch but, not knowing E's type, W would then end up paying that amount to all types of E including those that are indifferent between X and Y. By contrast, offering $t = 0$, which under the stated parameter condition is optimal for W, achieves efficient adjustment to Y when E is indifferent but not if $b = (b_H, b_L)$. This means that if $A = \{X, Y\}$, the parties will always switch from Y to X if X is more efficient, but if $A = \{X\}$, they cannot always switch from X to Y if Y is more efficient.

To conclude this section, E and W obviously need to agree on a set of tasks A simply to define the scope of W's obligation and to determine his wage. Unclear, however, is what criteria lead should to the exclusion of a task. Proposition 8 showed that including a high-cost (but sometimes efficient) task is optimal if the wage can be contingent on the task. Proposition 9 showed that for equal valuation supports, including a task is optimal whenever it is *sometimes* efficient. In practice, deliberately excluding tasks can be warranted if wages cannot be fine-tuned to actions, if negotiating spot adjustments is difficult, or (stepping outside of the model), the parties cannot agree on the payoff distributions of an action.

8 Multiple sellers

This section aims to shed light on why employment became a dominant way to work during the Industrial Revolution of the 18th and 19th centuries, and why it may be losing its dominance in the 21st century. To help explain the former, I show that complementarity in production across workers favors employment. To help explain the latter, I show that

ex-post competition among workers strongly favors spot trade.

8.1 Complementarity in production

Employment outside of a context of team production has always existed.[29] Nevertheless, the increase in wage labor during the 18th century appears to correlate with a greater reliance on the coordinated division of labor (McKendrick 1961, Millward 1981). I show that complementarity in production in firms creates externalities in the contracting relationships with workers, both with spot trade and with employment, which under plausible conditions favor employment.

Suppose now that there are two workers W1 and W2 whose work is complementary in the sense that successful production of a good requires both workers to provide work. Recall that in the binary model of Section 4, W's work ($a = 1$) or no work ($a = 0$) leads to a benefit ab for E and a cost ac for W. Now, instead, suppose that each of two workers either works or doesn't ($a_i \in \{0, 1\}$), and that E's resulting benefit *per worker* is bY with $Y = a_1 a_2$, with the supports and distributions of b and c being the same as in Section 4. The workers' costs are independent.

With spot trade, E makes simultaneous price offers $t = (t_1, t_2)$ to the two workers, who each accept or reject. To avoid stacking the cards against spot trade, assume that if, say, W1 accepts but W2 rejects and E thus knows that production cannot succeed, E can rescind the offer to W1.

Proposition 10. *With spot trade, for any given parameter values of the binary model, E's average price per worker is at least as high with complementary as with independent production, and her profit per worker is the same or lower.*

Proof: see the Appendix. "Average price per worker" refers to the fact that E may wish to offer the two workers different prices. The proposition is intuitive: when E knows that both workers must accept their prices for production to be feasible, she offers higher prices on average and ends up paying higher rents to the workers, or she offers the same prices but the workers are less likely to both accept.

If E instead employs both workers, both are required to work. That is inefficient if $b = b_L$ and $(c_1, c_2) = (c_{H,} c_H)$, but the workers can achieve an adjustment even without

[29] Examples include domestic servants and farm workers (Thompson 1967). A modern-day example is that of a dental hygienist, who obviously needs to coordinate work closely with their dental practice and its schedule, but whose actual work with each patient is solitary.

coordinating: a type-c_L worker can offer $t = b_L$ to E to switch to no work. Only b_L accepts the offer and only if both sellers make it, and efficiency is restored. The resulting expected gain per worker is lower than with independent production because this state occurs only with probability $(1-\gamma)^2 < 1-\gamma$.

If $b = b_L$ and $(c_1, c_2) = (c_H, c_L)$ or (c_L, c_H), work may or may not be efficient. In the context of factory work, for instance, it is plausible that even when production is less valuable, it is still efficient to produce if one of the workers (but not both) has high cost; i.e. if $2b_L > c_H + c_L$. Employment has a clear advantage in this case.

However, if $2b_L < c_H + c_L$, then E's order for the workers to work is inefficient, and adjusting to no work may fail because E needs to be compensated for $2b_L$ to agree, but only worker c_H is willing to pay. Switching to no work would succeed only if b_L is small enough for a worker with high cost to offer $t = 2b_L$ instead of $t = b_L$. The next result summarizes this discussion.

Proposition 11. *If $2b_L \geq c_H + c_L$, then employment is strictly more profitable per worker with complementary than with independent production. If $2b_L < c_H + c_L$, then employment can be less or more profitable than with independent production.*

Proof: see the Appendix. To conclude this section, if production requires the work of multiple people, spot trade is unambiguously less profitable than if workers work independently. If in addition production is valuable enough even if some workers have a high cost, then employment is also strictly more profitable than with independent production. However, since per our previous results team production is not a requirement for employment, the results support Williamson's (1975) response to Alchian and Demsetz (1972) that team production may contribute to what institutions we observe, but is not the underlying cause.[30]

8.2 Ex-post competition among sellers

The analysis so far has assumed that following ex-ante competition among workers, the creation of quasi-rents leads to a bilateral monopoly. Gig economy companies such as Uber, TaskRabbit, Upwork and others, however, have created platforms where for any

[30] "Our assessment of the technological nonseparability argument thus comes down to this: Such conditions are merely symptomatic of a set of underlying transactional factors which, both here and elsewhere, ultimately explain the organization of economic activity as between markets and hierarchies" (p. 61).

service requested by a customer, there are multiple workers available to do the task. In these cases, then, there is no fundamental transformation.

Consider a version of the binary model of Section 4 where employment works the same way as before, but with a spot market, there are n sellers available to perform a given task, at i.i.d. costs $c \in \{c_L, c_H\}$. A buyer with $b = b_L$ will then offer $t = c_L$, which is accepted by one of the sellers with probability $1 - (1-\gamma)^n$. If $b = b_H$, B can offer $t = c_H$ and earn payoff $b_H - c_H$. Or she can offer $t = c_L$, which again is accepted by one of the sellers with probability $1 - (1-\gamma)^n$, leading to payoff $[1 - (1-\gamma)^n](b_H - c_L)$, which for any γ exceeds $b_H - c_H$ if n is large enough. Clearly, the more sellers compete, the higher the payoff from spot trade, and we obtain

Proposition 12. *For any $\gamma \in [0,1]$ there exists $n \geq 1$ such that E's profit from spot trade with n competing sellers exceeds the profit with employment.*

Proof: see the Appendix. Recall that with employment, E's profit falls short of the first-best surplus because W earns both an obedience wage premium and an adjustment rent. With spot trade, however, as n grows, not only does the maximal surplus itself increase because $\min\{c_i\}$ converges to c_L, but E also in the limit captures all of the surplus.

The result highlights the importance that "small numbers" (Williamson, 1975) play for employment as an equilibrium governance structure.[31] If firms could each day fill their positions from a large pool of available workers, employment would not exist. It exists because turnover is costly for both sides, and for many reasons, which favors at least somewhat durable relationships between firms and their workers. This is where employment tends to be both more efficient and more profitable than the market.

Information technology, in particular, has helped reduce the specific investments that give rise to small numbers, and enabled ex-post competition for certain services that used to require hiring employees. Technology has made it possible to (i) find people to perform ad-hoc tasks, such as those offered by TaskRabbit, (ii) for workers to perform many tasks remotely, such as those available through Amazon Mechanical Turk, and (iii) to help overcome adverse selection and moral hazard with the help of ratings systems.

[31] "It is the thesis of this chapter that task idiosyncrasies are common, that these give rise to small-numbers exchange conditions, and that market contracting is supplanted by an employment relation pricipally for this reason." (Williamson 1975: 62)

9 Support from the 18^{th} and 21^{st} centuries

I have made two main claims: (1) Employment is a governance structure designed to ensure a predictable supply of labor. Team production, centralized ownership of productive assets, and the need to adapt work to changing business needs, all favor employment over outsourced labor, but none are essential to its nature. (2) Observed employment may or may not be efficient when workers lack wealth and compete for opportunities to work. In this section I support these claims, via several sub-claims, with observations from business history or with recent empirical or anecdotal evidence.[32]

1. Contract labor is the main alternative to employment.

In my model, the alternative to employment is market trade for labor, and I argued in Section 2 that the absence of trade is a limitation of several models of adaptation or authority. Contracting for labor was common in 18th century Europe. Under the putting-out system, "merchant-manufacturers 'put out' raw materials...to dispersed cottage labor, to be worked up into finished or semi-finished product." (Landes 1966, cited in Williamson 1985: 215).

Although the domestic producers sold services not products, they determined their own work hours and were paid by the piece.[33] "[The weaver] supplied tools and...was moreover not bound to a single master, ...[and] was naturally inclined to consider himself not as a workman, but as a contractor" (Mantoux: 64). Nevertheless, like some of today's gig workers, the producers often depended economically on the merchants: "[The merchant clothier] owned the raw material and consequently the product; those through whose hands this product passed ...were no more, in spite of their apparent independence, than workmen in the service of an employer" (Mantoux 1928: 62).[34]

The putting-out system was eventually replaced with factory production in most industries, but continued to exist in some industries well into the 19th and 20th centuries

[32] For other work drawing parallels between the early stages of industrialization and today, see Finkin (2015) and Juhász et al. (2020b).

[33] "A putting-out system was not an organization. It rather was a complex network of contracts of manufacture. The producers were formally independent, though often highly dependent economically" (Kieser 1994).

[34] Similarly, according to Landes (1966), "most domestic weavers...were not, however, inpedendent entrepreneurs selling their products in the open market; rather they were hirelings, generally tied to a particular employer, to whom they agreed to furnish a given amount of work at a price stipulated in advance" (cited in Williamson 1985: 216). See also the very detailed analysis of the putting-out system in Millward (1981).

(Finkin 2015). It even exists today, for instance in the knitware industry of Modena, Italy (Kieser 1994, Lazerson 1995).

Work in the 20th century was almost synonymous with employment for most anyone other than farmers and entrepreneurs.[35] Mirroring the 18th century, however, the 21st century has seen a rapid increase in the number of workers who sell their labor as contractors. Katz and Krueger (2019) estimate that in 2015, as many as 16% of American workers were engaged in "alternative work arrangements". Oyer (2016), counting secondary jobs as well, puts the share of "independent workers" at over 30%. Both studies report a rapid growth over the past decade. In a study of resumes on Indeed.com, "freelance jobs listed on the resumes we examined increased by over 500 percent" between 2000 and 2014 (Paychex 2016).

One key driver of this trend has been information technology, which allows many tasks to be performed remotely, and enables buyers and sellers to match with one another on platforms such as Upwork, Fiverr, or Toptotal.[36] Matching on these platforms isn't necessarily anonymous: similar to sequential spot contracting in my model, buyers on Upwork can request to hire the same person they hired for a previous job.

My model predicts that employment is profitable only if the quasi-rent between E and W is large enough. The quasi-rent depends on parameters such as the probability of high demand β, the frequency of interaction δ, and E's ability to bundle similar tasks into one job, all of which are correlated with firm size. Consistent with this prediction, "80% of buyers using Upwork's website are businesses with 10 or fewer employees" (Oyer 2016).

2. Employment increased along with firms' demand for a predictable supply of labor.

The increase in 18th century employment coincided with the concentration of production in factories. The use of machines, which in England became widespread around 1760, was merely the last stage of this process, and made concentration inevitable due to the machines' size, scale, need for power, and cost (Landes 1986: 606, see also Juhász et al

[35] For evidence, see Parsons (1986). Or as Simon (1991) put it, "no matter whether our visitor [from Mars] approached the United States or the Soviet Union, urban China or the European Community, the greater part of the space below it would be within the green areas, for almost all of the inhabitants would be employees, hence inside the firm boundaries."

[36] Consumer-focused platforms such as TaskRabbit or Handy, in turn, enable workers to bypass agencies (such as residential cleaning services) as middlemen. Here, the employer in question is not the actual labor buyer (the consumer) but the agency.

2020a).

Well before that, however (even starting in the 16th century), production in many industries gradually shifted from workers' homes to factories. "The manufacturer ... kept the tools, and organized workshops under his direct supervision, whilst the artisan sold him only his labour, for which he received a wage" (Mantoux: 36). Physical concentration permitted a greater division of labor and the associated productivity gains than the putting-out system did, as illustrated by the low-tech pin factory that Adam Smith described.[37] [38]

Some factories even used the same technology as the home industries they replaced, which suggests that organizational efficiency and not just technological change led to the concentration of production. The old putting-out system was

> "plagued by problems of irregularity of production, loss of materials in transit and through embezzlement, slowness of manufacture, lack of uniformity and uncertainty of the quality of the product. But most of all, they were limited by their inability to change the processes of production (Braverman 1974, cited in Williamson 1985: 232)."

Accordingly, as markets grew and competition increased, "considerable economies in costs could be achieved purely by centralising production without technical change" (Hudson 1981, see also Clark 1994). While Williamson (1985) and others emphasized efficiency gains, Marglin (1974) concluded from much the same facts that "rather than providing more output for the same inputs, these innovations in work organization were introduced so that the capitalist got himself a larger share of the pie at the expense of the worker."

In my model, in which technology is fixed too, both effects can be at work. An increase in demand for the quantity and predictability of output can be interpreted as as increase in the entrepreneur's benefit b, which according to Propositions 3 and 5 makes employment

[37] An equally low-tech example from literature is the factory of Jean Valjean, the protagonist of Victor Hugo's *Les Miserables*. The factory makes "cheap black beads" using simple technology "that could have been carved out at almost any time, since the raw materials, the process and the product were all available before 1816" (Bellos 2017: 69-70).

[38] As Stigler (1951) cautioned, though, physical proximity permits a greater division of labor even in the absence of hierarchy. In the small-arms industry of Birmingham around 1860, for instance, "specialism was carried out to an almost unbelievable extent" within an ecosystem of thousands of craftsmen in dozens of specializations that combined both genuine market trade and putting-out, see Allen (1929).

relatively more efficient. The profit difference ΔV, however, consists of both efficiency gains ($D^0 - D^E$) and E's better ability to extract rents from W under employment (the expression $U^0 - (1-\delta)U^E$ discussed in the context of Proposition 2). Thus, in a long-standing debate between mainstream and radical economists about the origins of capitalist hierarchy (see Williamson 1985, chapter 9), both sides may be right.

In the 21st century, the link between employment and the physical concentration of work is weakening because more and more employees can work from home, a trend that has likely been permanently accelerated by the Covid-19 pandemic (Barrero et al. 2021). Information technology is again a key driver, by facilitating communication and performance measurement. What distinguishes today's remote employee from the freelancer is neither the workplace (often, home) nor the productive assets used (say, a computer), but the employee's obligation to work on terms defined by the labor buyer and, in return, his regular paycheck.

3. Historically, many employees have had only a single task.

Almost all economic theorizing about employment emphasizes the adaptability that it affords. In apparent contrast, however, extreme specialization and rigor characterized the work of countless employees from the 18th century until well into the 20th century, and still exists today. As I have argued, this can be explained if employment is more generally thought of as a relational obligation to work, irrespective of how broad or narrow the worker's responsibilities are.

A well-documented and large-scale example of 18th century labor specialization can be found in Josiah Wedgwood's pottery works, where "his workmen ... were trained to one particular task and they had to stick to it. ... Out of the 278 men, women and children that Wedgwood employed in June 1790, [all but five] were specialists." (McKendrick 1961).[39]

The trend toward specialization and rigor, rather than flexibility and adaptation, only became more pronounced when machines were more widely used:

> "the manufacturer ... had, so to speak, to turn [all these unskilled men] into a human machine, as regular in its working, as accurate in its movements, and

[39] "Each workshop had its specialists: in coloured ware, there were painters, grinders, printers, liners, borderers, burnishers and scourers; in jasper, there were ornamenters, turners, slip-makers, grinders, scourers, and mould-makers; in black ware, there were turners, throwers, handlers, seal-makers, mould-makers and slip-makers; and in all there were modellers, firemen, overlookers, porters and packers."

> as exactly combined for a single purpose, as the mechanism of wood and metal to which they became accessory. Hard-and-fast rules replaced the freedom of the small workshops. ... Within the factory each had his allotted place and his strictly defined and invariable duty. " (Mantoux: 384)

The specialization of work may have reached its peak in the early 20th century with the arrival of mass production, combined with the application of Frederick Taylor's Scientific Management (Lindbeck and Snower 1996, Hu 2013, Kranzberg and Hannan 2017). Assembly-line work was pioneered by Ford in the 1910s, where "the extreme specialization of the work obliges the man to make the same motion over and over again" (Ford Man 1917, cited in Lewchuk 1993), and its monotony was immortalized 20 years later in the factory scene of Charlie Chaplin's *Modern Times*.

Similar conditions existed well into the late 20th century, when for instance in 1972, workers revolted against alienating job conditions at GM's factory in Lordstown, Ohio: "The young lords of Lordstown found the assembly line — 35 second bursts of a dull, repetitive task, and a 5-second break before the next Impala or Vega rolled up — to be soul-crushing work." (Bunch 2018). Efforts to reorganize work from "tayloristic" to "holistic" (Lindbeck and Snower 1996, 2000), including a wider variety of tasks, only gained traction around that time, i.e. during the last 50 of over 250 years of industrial-age wage labor.

4. 18th and 19th century workers resented and resisted the rigors and control of employment, thus displaying a strong preference for control over their time.

In the abstraction of social science, employment is a contract whereby a worker trades control over his actions for a wage. In reality, this "contract" was a hard-fought struggle between workers and capitalists that spanned most of the 18th and 19th centuries, until finally employment became so pervasive that its expectations were fully ingrained in societal norms. Workers' aversion to the loss of autonomy can be interpreted in psychological terms (Stone et al. 2009). More tangibly, though, it also reflects workers' preference for control of their time, which enter in my model in the form of stochastic opportunity costs of working.

Pollard (1963) argued that "one of the most critical, and one of the most difficult,

transformations required in an industrializing society is the adjustment of labour to the regularity and discipline of factory work". Mantoux (p. 419) elaborates:

> "The feeling of repulsion which [the factory] aroused is easily understood, as, to a man used to working at home, or in a small workshop, factory discipline was intolerable. Even though at home he had to work long hours to make up for the lowness of his wage, yet he could begin and stop at will, and without regular hours. He could divide up the work as he chose, come and go, rest for a moment, and even, if he chose, be idle for days together. ... He was not bound by hard and fast regulations, as relentless and as devoid of sympathy as the machinery itself."

"As a result of this attitude, attendance was irregular, and the complaint of Edward Cave, in the very earliest days of industrialization, was later re-echoed by many others: 'I have not half my people come to work to-day, and I have no great fascination in the prospect I have to put myself in the power of such people."' (Pollard 1963).[40]

Accordingly, the most successful industrialists were neither merchants nor inventors, but those who succeeded at *organizing* their factories:

> "The main difficulty ... did not ... lie so much in the invention of a proper self-acting mechanism for drawing out and twisting cotton into a continuous thread, as in...training human beings to renounce their desultory habits of work, and to identify themselves with the unvarying regularity of the complex automation. To devise and administer a successful code of factory discipline, suited to the necessities of factory diligence, was the Herculean enterprise, the noble achievement of [Richard]Arkwright," a pioneer of 18th century industrialization (Ure 1835: 15).[41]

Workers' preference for control over their work hours is asserting itself again today, and is a second key driver of the rise of contract labor in the 21st century. "The IW

[40] Marglin (1974) adds: "Discipline did not improve once the Cave factory became mechanized. When Wyatt visited the new spinning mill at Northampton in 1743 he found that 'only four frames were regularly at work, since there were seldom hands enough for five."' Still in the 18th century but 50 years later, "It was a constant test of Wedgwood's ingenuity to enforce six hours of punctual and constant attendance upon his workers, to get them to avoid waste, and to keep them from drinking on the job and taking unauthorized 'holidays"' (Kranzberg and Hannan 2017).

[41] In addition to human-resource challenges, early industrialists had to learn, by trial and error, how to organize factories from an operations perspective, see Chapman (1974) and Juhász et al (2020a).

[independent workers] set their own hours, both in terms of overall quantity and the specific times they work. This allows them to balance work, family, and leisure activities better than traditional employees" (Oyer 2016).

Chen et al. (2019) estimate Uber drivers' opportunity costs of working over short time intervals, and find large fluctuations, suggesting a high value for those drivers' control over their time. My model predicts precisely this kind of self-selection between employment and gig-worker status: where workers compete to be employees, those with smaller fluctuations in the opportunity costs of their time will accept a lower obedience wage premium and will underbid workers whose costs vary more strongly.

5. Employment and the lack of ownership of productive assets often go hand in hand, but are not the same thing.

Property-rights theory and incentive-system theory tell us that productive assets should be owned by those best positioned to earn a return on them, and that control over non-human assets confers influence over workers (Holmström and Milgrom 1994, Hart 1996, Holmström 1999). Both theories thus help explain the high correlation between employment and the lack of asset ownership. However, they do not distinguish between market and hierarchy, and thus gloss over unique features of firms that Coase, Simon, and Williamson sought to explain. *Defining* employment by the lack of asset ownership therefore seems unsatisfactory. In addition, several examples illustrate that the connection between asset ownership and employment is not always tight or simple:

1. Ownership of productive assets need not confer ownership of the product. Under the putting-out system, the merchant owned the output because he owned the raw material; he purchased the raw material because the quality of the final product significantly depended on it. The worker in his home owned the tools of production, which gave him optimal incentives to maintain them. Though technically a contractor, the worker's lack of ownership of the output nevertheless often led to a high degree of economic dependence. If ownership of the product is considered the key asset, then property-rights theory would classify workers as employees, contrary to the consensus of historians.

2. Contractors often do not own any relevant non-human assets.

(i) The workers of the putting-out system were still contractors even when, due to poverty, they lost ownership of their tools to the merchants, or were never owners to begin with. "The frame-work stocking-knitters in London and Nottingham paid rent—frame-rent—for the use of their knitting frames" (Mantoux: 65) but remained in full control of when they worked.

(ii) Under the inside-contracting system that was common in New England between the Civil War and World War I, capitalists supplied all equipment and raw materials, and sold the final product. "The gap between raw material and finished product, however, was filled not by paid employees arranged in the descending hierarchy so dear to the hearts of personnel experts but by contractors, to whom the production job was delegated" (Buttrick 1952, see also Williamson 1975, 1985).

(iii) In the 21st century, inside contracting can take the form of "independent contractors taking on managerial roles in client organizations where they manage and direct the employees of their client" (Anderson and Cappelli 2021), and some startups even engage CFOs as contractors (Broughton 2021).

3. Employees sometimes own their tools, especially when they require good incentives for maintenance but are portable enough to be in the possession of the workers. Thus, most orchestra musicians own their instruments; auto mechanics have their own hand tools; so do carpenters and hairdressers. All of these professions have both employees and contractors, whose status is not primarily determined by asset ownership.

4. For a worker who switches his status from contractor to employee or vice versa while working for the same firm, as happens on occasion in many professions, many things change such as schedules and method of pay, but a change in asset ownership on either side is usually not involved.

Van den Steen (2010) shows that a worker's lack of assets and therefore of outside options supports an authority relationship. In my model, this link appears in reduced form in the distribution of b and c: Shifting a productive asset from W to E plausibly lowers c (by lowering W's outside options) and raises b (by raising E's demand for labor to complement the asset's productive capacity), which per Propositions 3 and 5 raises the absolute and relative profitability of employment. Then again, in my model

productive assets matter for E's employee-or-contractor decision only inasmuch they affect the benefits and costs of a worker's labor. Because these are the fundamentals of the problem, assets need in fact not play any role, as is arguably the case with any work requiring only a PC and widely available software.

6. Poverty drove many workers into low-wage employment, which may have eroded the competitiveness of independent labor.

Factories with employed workforces eventually came to dominate production because they were both technologically and organizationally most efficient. At least in England, however, the transition was also facilitated by "the emergence of an increasingly landless and wage-dependent rural proletariat" (Hudson 1981, see also Mantoux: 74). The domestic producers were independent but poor, and their poverty caused many to eventually lose all of their productive assets.

> "If it was a bad year and the harvest was deficient, ...he had to borrow, and who was the most likely person to lend if not the merchant who employed him? The merchant was generally willing to lend him money, but he needed security, and the readiest pledge was the weaver's loom which, after becoming the means of earning mere wages, now ceased to be the exclusive property of the producer. In this way, following on the raw material, the implement in its turn fell into the capitalist's hands. ...Thus the producer, gradually deprived of all rights of ownership over the instruments of production, had in the end only his labor to sell and his wages to live on." (Mantoux: 64-65).[42] [43]

Landlessness, in turn, was for many peasants the result of the Enclosure Acts and drove them into cities, again with only their labor to sell in order to make a living. This state of affairs represents an extreme version of $\tau = 0$ in my model, where workers without wealth compete to do business with the entrepreneur.

[42] Mantoux continues: "His position was even more precarious when, instead of living in the country, where the land itself still helped him to make a living, he lived in the town inhabited by the merchant clothier. Then he became completely dependent, having none but the clothier to look to for the work on which he lived" (p.65).

[43] Or as Marx famously concluded: "He, who before was the money-owner, now strides in front as capitalist; the possessor of labour-power follows as his labourer. The one with an air of importance, smirking, intent on business; the other, timid and holding back, like one who is bringing his own hide to market and has nothing to expect but – a hiding" (1867, p.123).

As Example 1(c) in Section 4 suggests, then, it is possible that in some cases employment was less efficient but more profitable than contracting (i.e., putting-out). Employment could have been less efficient if (before the availability of machines) no change in technology was involved, due to the managerial hurdles of organizing reluctant workers into factories. Employment could have been more profitable, though, because it better enabled merchant-manufacturers to extract rents from workers; cf. Proposition 2. In that case, even those domestic producers who retained ownership of their tools could have been underbid by workers willing to accept a low wage for employment.

10 The legal status of employment

I have followed Coase and Simon in depicting employment as a contract. The added nuance that such a contract is best understood to be relational helps explain how authority can exist in a market economy governed by contract law, even when "the subordinate cannot transfer the actual performance of the action" (Coleman 1990: 79). Nevertheless, employment and market transactions are in fact governed by different laws; see Masten (1988) and Williamson (1991). Especially in light of much recent controversy over the legal status of gig economy workers (see Posner 2020), it seems useful to clarify the role of the legal environment. In addition, this paper's analysis may help to inform the current normative legal debate.

Two arguments justify a contractual, as opposed to a legal, approach to studying employment. One is historical: Employment is too pervasive an institution across time and different places for any particular legal regime to have much explanatory power. Also, and again across time and different places, legal regimes usually evolve in response to economic circumstances, not the other way around (for instance, on constantly evolving laws in Britain throughout the 18th and 19th centuries, see Mantoux, 1928). The bottom line is that although all countries today have their employment laws, employment as economic institution probably does not owe its existence to any of them.

Second and more relevant to current business practice, at least in the Unites States and in other countries, a firm and a worker cannot simply choose a legal regime—contract law or employment law—to organize their relationship. Instead, a worker's legal status is determined by the economic nature of the relationship between firm and worker, not by whether the firm calls the worker "employee" or "contractor" (see, for example, Internal

Revenue Service 2017).[44]

At stake is that contractors are entitled neither to minimum wages nor to benefits such as unemployment insurance or employer contributions to social security. This may create an incentive for firms to label their workers as contractors even if economically, their work resembles employment. Misclassification claims by workers can trigger audits and possibly fines by the IRS or state or federal Departments of Labor, or play out in court.

These authorities adjudicate a worker's employment status based on criteria such as behavioral control, financial control and dependency, ownership of assets used by the worker, centrality of the worker's job to the firm's business, and the duration and exclusivity of the relationship.[45] Although "pure" employment and "pure" independent contracting are common enough, many real-life work relationships fall—like the putting-out system in the 18th century did—on a continuum between them, which is where the designation can become ambiguous and controversial.

The growth of the gig economy and of alternative work arrangements has led to a proliferation of legal cases against firms, as well as to intense debate in the legal profession over the normative case for distinguishing employees and contractors (see Posner 2020). Among legal cases, a prominent example is Uber. In April 2019, the federal Department of Labor issued an opinion supporting Uber's treatment of drivers as contractors:

> "Drivers' virtually complete control of their cars, work schedules, and log-in locations, together with their freedom to work for competitors of Uber, provided them with significant entrepreneurial opportunity. On any given day, at any free moment, UberX drivers could decide how best to serve their

[44] This argument contradicts Williamson (1991) as far as *choice* is concerned:

> "There is a logic to classical market contracting and there is a logic for forbearance law, and the choice of one regime precludes the other. Whether a transaction is organized as make or buy—internal procurement or market procurement, respectively—thus matters greatly in dispute-resolution respects: the courts will hear disputes of the one kind and will refuse to be drawn into the resolution of disputes of the other."

Instead, as Williamson added as caveat, "Personnel disputes are more complicated": Courts will in fact hear disputes over whether a (personnel) transaction is "make" or "buy" in the first place.

[45] To give two examples, the IRS uses 11 criteria in three groups: behavioral control, financial control, and features of the relationship. (Internal Revenue Service 2017). California's Assembly Bill 5 passed in 2019 mandates an "ABC test," according to which the most stringent requirement for contractor status is "The worker performs work that is outside the usual course of the hiring entity's business" (Department of Industrial Relations 2019).

> economic objectives: by fulfilling ride requests through the App, working for a competing rideshare service, or pursuing a different venture altogether" (National Labor Relations Board 2019).

A San Francisco court, however, ruled in August 2020 that according to California's more stringent criteria, Uber drivers are employees because their services are central to Uber's business (Siddiqui 2020). Likewise, a London tribunal opined in 2016: "The notion that Uber in London is a mosaic of 30,000 small businesses linked by a common 'platform' is to our minds faintly ridiculous" (cited in Cunningham-Parmeter 2019).

My model focuses on behavioral control, and defines the distinction between employee and contractor as the presence or absence of the worker's obligation to work for a firm. According to that distinction, Uber drivers are unambiguously contractors, based on the description of their job in the DOL's opinion quoted above.

Other dimensions of behavioral control seem more questionable as criteria for telling apart employees and contractors. Both federal and state authorities, for instance, take into account how much the *execution* of a task is subject to control by the labor buyer. But control over execution seems mostly driven by the buyer's business needs and the expertise of both parties: I won't tell a plumber how to fix my faucet, but I may specify in minute detail how MTurk contractors are to collect data for me. Likewise, Amazon's coders and its warehouse workers have vastly different degrees of autonomy, but both are employees.

Relatedly, this paper's contractual approach runs counter to the most controversial aspect of the "ABC test" used in many U.S. states, according to which a worker cannot be classified as contractor unless "the worker performs work that is outside the usual course of the hiring entity's business" (Department of Industrial Relations 2019). While businesses may well prefer to hire employees instead of relying on contractors especially when team production is involved (see Propositions 10 and 11), there is no good economic reason to *require* firms to treat workers as employees only because they perform business-central tasks.[46]

A second distinction drives the contrast between Proposition 12 and all other results of

[46] As mentioned, it is this part of the ABC test that led to California's designation of Uber drivers as employees. Proposition 22 was a ballot initiative introduced in the November 2020 state election to grant app-based transportation and delivery companies an exemption from California's statute. It passed with 59% of the vote, including overwhelming support by Uber drivers.

the paper: the presence or absence of a Fundamental Transformation that creates quasi-rents between firm and worker, resulting from relationship-specific investments. Relatedly, Posner (2020) recently suggested that workers should be classified as employees, and thus enjoy the protections of employment and labor laws, if due to relationship-specific investments they are subject to a firm's monopsony power.

In spite of some overlap in spirit between Posner's discussion and my model, relationship-specific investments play different roles in both papers. Posner argues that subsequent to making specific investments, workers face high exit costs, which cause them to submit themselves to the employer's control. In my model, instead (and as Williamson 1975 argued), *the entrepreneur's* cost of switching workers is the driving force that creates a risk of holdup by the contractor, and that makes employment more efficient and/or more profitable by requiring the worker to provide his service.[47]

11 Beyond employment: a simple theory of the firm

The difference between authority and the price system is also central to firms' vertical integration decisions, and Williamson (1975: 82) argued that the key tradeoffs in firms' make-or-buy decisions are the same that arise when hiring individual workers. They are simpler, in fact, in that obedience is less of a problem. If a buyer of labor employs the seller (the worker), her authority is not legally enforceable. If the seller is a firm, however, then a buyer that integrates the seller acquires the legal right to make decisions pertaining to the seller's non-human assets.

Contractible authority is a special case of the above analysis for δ set to 1. The model of bilateral trade developed above therefore provides a simple but fully micro-founded theory of the firm that puts market and hierarchy on a level playing field. Its main limitation is the absence of incentives for effort or investments.

Note first that the model generates, under fully standard assumptions, a key prediction of transaction-cost theory "that any increase in quasi-rents will increase the likelihood of vertical integration (a finding that is so far consistent with nearly all of the existing empirical literature)" (Whinston 2003); see Propositions 3 and 5. This prediction follows from the benefit of shifting the disagreement point under incomplete-information bargaining

[47] Relatedly, Posner emphasizes that "economic dependence" should be understood to mean exposure to monopsony, not poverty. Here, however, the worker's lack of wealth matters too.

to the party that makes more efficient decisions. By contrast, as Holmström and Roberts (1998), Whinston (2003), and Gibbons (2005) discuss, transaction-cost models tend to *assume* that adaptation is more efficient within firms (or equivalently, that haggling over rents matters less), whereas property-rights theory makes no prediction pertaining to levels of quasi-rents.[48]

Within Gibbons' (2005) taxonomy, my model combines the adaptation and rent-seeking perspectives, which are both central to transaction-cost theory. The former focuses on the relative ability of firms and markets to tailor actions to the state of the world; the latter focuses on socially inefficient haggling over quasi-rents. As Gibbons notes, both approaches focus on ex-post actions rather than ex-ante investments. Both types of models, however, tend to rule out market trade.

In my model, adaptation and rent seeking are two sides of the same coin: adaptation is the goal and rent-seeking the cost, both with spot trade and with integration. Microfounded by incomplete information, destructive haggling and seller opportunism are indeed the costs of the market here—and they are high especially if trade is valuable but no trade is the default. The theory makes precise how the firm can do better, namely by shifting decision rights to the party that on average makes more efficient decisions.

Especially because firms tend to be observed when they are more efficient than markets (Gibbons 2005), the "supersession of the price mechanism" (Coase) or "fiat vs. haggling" (Williamson) are then merely symptoms rather than defining distinctions between firm and market: In the market, we observe haggling because without it, trade doesn't occur. Within firms, if the decision-right holder's default decision is efficient or close to efficient, the "fiat" sticks when, because of incomplete information, any further gains from trade are too small for intra-firm haggling to change the outcome. That is, in my model the absence or rarity of intra-firm bargaining is an equilibrium outcome, not a design feature.

Relatedly, Williamson's (1973) intuition that "substantially the same factors that are ultimately responsible for market failures also explain failures of internal organization" certainly holds in my model, in the sense that the underlying cause of failure in both cases is an inefficient default outcome determined by the party in control. In the market, if is trade is valuable, there will be haggling between opportunistic agents to escape the

[48] On TCE, see Masten (1986), Tadelis (2002), and Tadelis and Williamson (2013, Assumption 1). As the latter note, most assumptions of TCE have various microfoundations, but "a formal integration of these micro-foundations has not been performed" (Tadelis 2002).

no-trade default. In the firm, a boss's inefficient order either amounts to abuse of fiat (Dow 1987, Gibbons 2010) or leads to influence activities by others to change the outcome (Milgrom and Roberts 1988, 1990; Gibbons 2005, Powell 2015). Though the symptoms of failure may appear different, therefore, their cause is the same.

Aside from allowing for market transactions, the bargaining framework differs from rent-seeking models in that not all haggling is wasteful; some of it shifts surplus from one party to another. With unconstrained upfront transfers, those shifts don't matter for equilibrium outcomes. When upfront transfers are infeasible due to wealth constraints or any other reason (Williamson 1975: 67-68), however, rent extraction matters as much as deadweight loss, as can be seen in equation (7): it does not matter to E whether the cost of the market is failure to trade (D^0) or the seller's profit (U^0); all that matters is that E cannot capture the first-best surplus.

A key formalization of transaction-cost economics is Bajari and Tadelis (2001, henceforth BT in this paragraph), who also model adaptation through ex-post bargaining under incomplete information. Interpreting BT as a theory of the firm (as Tadelis 2002 does), BT and my paper emphasize different advantages of integration: BT focus on incentive alignment (the payoff function) and my paper on control rights (the disagreement point). That is, in BT an integrated seller retains his decision rights but his incentives are partially aligned with the buyer. This alignment makes intra-firm bargaining more efficient than bargaining across firms under high-powered incentives. In my model, in contrast, incentives are held fixed (formally, the distribution of b and c does not depend on the governance structure) but integration shifts decision rights over the seller's operations to the buyer. Both mechanisms—the "adaptability advantages" of low-powered incentives, and fiat vs. haggling—are emphasized in Williamson's work. BT offers a microfoundation of the former and mine of the latter.

As an application of the theory, take General Motor's acquisition of Fisher Body in 1926, the subject of a sizable literature (see Gibbons 2005 for references). GM and Fisher's relationship prior to acquisition was governed by an incomplete long-term contract, which resembles the case of sequential spot trade in this paper, because specific investments made in the past had already created substantial quasi-rents, and changing market conditions required intermittent renegotiation. According to Gibbons (2005), "the formal contract between GM and Fisher is said to have encouraged Fisher to take socially inefficient actions (e.g., regarding plant location), so as to increase Fisher's profit, at disproportionate

expense to GM. To stop this hold-up, GM eventually acquired Fisher."

In this story, adaptation to GM's increased demand for cars is the goal and rent seeking the cost. Rent-seeking usually goes both ways, however. In my model, inefficiency results from *both parties'* attempts to secure rents in the course of incomplete-information bargaining. For instance, if the buyer has all bargaining power, inefficiency results if she is unwilling to pay enough for the seller's service.

A more recent example is the interaction between major airlines and regional feeders as studied by Januszewski Forbes and Lederman (2009, 2010). Bad weather at a hub airport may create a high value to the major airline for the regional airline to make adjustments. As independent firm, the regional airline may have high opportunity costs, or may try to take advantage of the major. A major that owns the regional airline can instead simply dictate the desired adjustment.[49] The discussion of Section 5.2 applies, however: the exercise of authority is not per se more efficient than the use of prices. Instead, authority is more efficient if the adjustments favored by the buyer are on average more efficient than no adjustments, because bargaining is biased towards the disagreement point.

The above analysis assumed that workers are wealth-constrained, as many are. For supplier firms, that may or may not be the case. Whether wealth constraints, when they are present, matter for vertical integration depends on the timing of events. They do not seem to matter in the GM-Fisher Body example, where whatever ex-ante competition among sellers there may have been lies in the irrelevant past, and bilateral monopoly is therefore the starting point. In this situation, if GM acquires Fisher, its owners need to be fully compensated for their rents earned with spot trade. It follows that an acquisition can take place only if it is efficient. This scenario seems common.

By contrast, profitable but inefficient integration can occur if the buyer's make-or-buy decision occurs sooner, namely if and when sellers compete ex-ante, and the sellers' market valuations do not include any monopoly rents. It may also occur if the buyer initially opted for "make" but a more efficient supplier appears. If switching to "buy" would lead to rents for the new seller for which he is unable to compensate the buyer upfront, the buyer may prefer to stick with "make" even if "buy" is more efficient.

[49] In this example, a benefit of outsourcing not captured by the model is the lower labor costs of (non-unionized) independent regional airlines (Januszewski Forbes and Lederman 2009).

12 Conclusion

This paper's main contribution is to develop a theory of employment that builds on, formalizes, and clarifies arguments made by Coase, Simon, and Williamson. At the same time, the theory absorbs Alchian and Demsetz's criticism of those arguments, which for thirty years in effect discredited the Coase-Simon view of employment, until the arrival of the research discussed in Section 2. Envisioning authority as being supported by a relational contract (as Bolton and Rajan [2003] first did) reconciles sociologists' notion of authority with Alchian and Demsetz's assertion that "the firm ... has no power of fiat."

Methodologically, the paper's contribution is to model employment around a framework of bargaining under two-sided incomplete information. This makes it possible to compare employment with firms' main alternative, spot trade for labor, within the same economic environment and under conventional assumptions. Other related work, in contrast, either considers non-integrated structures without trade, or models costly trade by relying on specific informational assumptions, behavioral assumptions, or reduced-form costs of bargaining.

The logic of employment in my model is closer to Coase (1960) than to Coase (1937). While Coase's 1937 paper argues that employment replaces costly bargaining with orders, the 1960 paper shifts the attention to control rights as the default outcomes against which the parties can engage in costly bargaining. How common intra-firm (within-employment) bargaining actually is in real life is immaterial; the point is that market and hierarchy differ primarily in their default outcomes defined by control rights, not in whether or not bargaining takes place. As discussed in Section 11, "fiat vs. haggling," "supersession of the price mechanism," "seller opportunism," "abuse of fiat," and "influence activities" all emerge endogenously from the allocation of default control rights.

My characterization of employment as a relational obligation to work is primarily a result, but it also serves as a useful generalization of Simon's (1951) definition. It is a result, first, because starting from Simon's definition, employment in a model with only one task amounts to nothing more than an obligation to work. Second, the key drivers of what makes employment efficient or profitable already all appear in the one-task model. Whether the worker has multiple tasks, or the entrepreneur has important information, or who owns which assets, all matter for surplus and for profits, but not for the nature of the relationship.

Focusing on the employee's obligation to work, instead of on the boss's right to direct the worker, significantly expands the domain of the theory relative to Simon's: It covers employees whose job entails only a single task; it covers firms for which a predictable supply of labor is more important than flexibility in its use; and it covers employees who enjoy a significant amount of autonomy but who are still expected to show up for work.

Finally, workers' wealth constraints matter in the model, as they probably do in real life too. At stake in the model is not whether workers have the money to own productive assets, but whether they have the money to make upfront payments to a labor buyer to compensate for any rents they might earn later. Williamson (1975: 69-70) considered such payments unrealistic, but—like organizational economics in general—focused on explaining economic institutions in terms of efficiency (Williamson 1985, Roberts 2004). However, just like efficiency and profit maximization often diverge in mechanism design, they may also diverge when it comes to markets and hierarchies. Also, a simple assumption such as a worker's lack of wealth can lead to predictions rather close to some claims of radical economists, notably that employment may be more profitable even if spot trade is more efficient—one economist's "rent extraction" is another's "exploitation."

The most obvious limitation of my analysis is the omission of effort; i.e., of a distinction between perfunctory performance ("job performance of a minimally acceptable sort") and consummate performance (an "affirmative job attitude"), in Williamson's (1975: 69) terms. This omission does not bias my results in either direction because effort matters both across and within firms, and because both contracts and commands are incomplete (see Cheung 1983, Simon 1991): "[C]ourts cannot ... determine whether workers put their energy and inventiveness into the job" (Williamson 1975: 69), just like managerial authority "only obligates employees to perform duties ... in accordance with minimum standards" (Blau 1964: 206).

Eliciting consummate performance likely relies on relational contracts both within and between firms: Provided that performance can be observed, the employee will not just show up but also put forth effort in order to keep his job; so will the contractor, in order to earn repeat business, as in Klein and Leffler (1981). Although only formal analysis can deliver precise answers, it seems likely that the distinction emphasized in this paper remains relevant: No matter how complex the job, the employee has a boss

who exercises authority in some form (Simon 1991), the contractor does not.[50] This would also explain why, precisely as Foss (2002) argued, the knowledge economy has not made authority obsolete, contrary to predictions by some organization theorists. That is, although the Coase-Simon view of authority in its simplest statement may evoke "Theory X" (McGregor 1957) work environments, its key ideas, appropriately amended, are no less relevant in the 21st century.

Finally, less amenable to formal analysis but no less important, workers' identification with organizational goals is an important driver of motivation, as Barnard and Simon already observed 80 years ago (see also Cassar and Meier 2018, Freeland and Zuckerman 2018). Identification with the firm and its mission matters to this paper's topic mainly because it requires group membership that possibly only an employee, not a contractor, can have. Firms' ability to harness identification as motivator then depends not, or not only, on how work is organized contractually (as studied in this paper) but on who perceives himself to be a member of the group; see Gartenberg and Zenger (2021).

Appendix: Proofs

Proof of Proposition 1: 1. I first show that an optimal contract will not be task-contingent. This follows from the fact that obedience is the only moral-hazard dimension of the relationship, and instances of disobedience by W are observable and are punished with the strongest possible punishment, termination. It follows that while the relationship continues, there is never a reason to sacrifice surplus. In addition, there is no conflict between E's profit maximization and efficiency. To see this, note that for surplus $s = b - c$ and tasks k and l,

$$E[s^k|b^k] \geq E[s^l|b^l] \iff b^k - E[c^k] \geq b^l - E[c^l] \iff b^k \geq b^l,$$

where the first equivalence follows from independence of p and q and the second follows from the equality of the marginal distributions p^k and p^l. It follows that if E chooses the task that maximizes her profit $b^k - w_\tau(k)$, she will *always* choose $\arg\max_k b^k =$

[50] In an opinion piece, Asghar (2013) argues that university presidents have a harder job than CEOs because CEOs can and do exercise authority, including by firing executives, to enforce their direction. University presidents have no such powers in dealing with the most powerful group of employees, tenured faculty.

$\arg\max_k s^k|b$ if and only if $w_t(k) = w_t$ for all k. Any contract with task-contingent $w_t(k)$ can therefore be weakly improved, in terms of E's orders, by switching to $w_\tau = E[w_\tau(k)] = \overline{w}$ for all k, where expectations are taken over b and E's induced orders. Picking this particular wage also preserves incentives in prior periods $\tau' < \tau$, which only depend on $E[w_\tau(k)]$ and not on events in t.

To complete this step we need to check W's obedience constraints. Given that the highest possible cost $c_H = \max C$ is the same for all tasks, the most restrictive obedience constraint for each task k is

$$(1-\delta)(w_\tau(k) - c_H) + \delta U_{\tau+1}(k) \geq 0,$$

allowing for the possibility that the continuation utility $U_{\tau+1}$ depends on the task chosen in τ. Now if $w_\tau(k) = \overline{w}_\tau$ for all k, then $U_{\tau+1}$ can be task-independent too and can be held to the smallest value that satisfies both current and future incentive and participation constraints. However, if $w(k) > \overline{w} > w(l)$ for some k and l, then while the lower wage $w(l)$ can be made up for by a higher wage later, the reverse is not true for task k: It may not be possible to counterbalance an above-average $w(k)$ in τ by a lower $U_{\tau+1}(k)$, if doing so violates constraints in $\tau+1$. It follows that in terms of W's obedience constraints, too, there is nothing to gain from differentiating the wage by the task ordered.

2. Next, I show show that a stationary contract is optimal as long as spot adjustments are not affected. Key to the result is that E's optimal order, and thus W's eventual action, does not depend on w. Therefore, any feasible wage profile $\{w_t\}$ that satisfies W's obedience constraint in every period,

$$(1-\delta)(w_t - c_H) + \delta \widetilde{U}_{t+1} \geq 0,$$

will lead to the same stochastic sequence of actions m_t^E. In particular, while E could, for any profile $\{w_t\}$, reduce w_1 and pay a higher wage later, doing so would only make later obedience constraints slack without creating any allocative benefit. Thus, any wage profile $\{w_t\}$ that satisfies

$$w_t - c_H + \sum_{\tau=1}^{\infty} \delta^\tau \left[w_{t+\tau} - c^B + U^E \right] \geq 0 \quad \text{for all } t \geq 1$$

can also be satisfied for t by the stationary contract

$$w = c_H - \delta \left[c_H - c^B + U^E \right],$$

and in that case it satisfies all future incentive constraints $t' > t$ as well.

3. Even though E's initial order $\widetilde{m}^E$ does not depends on w if it is constant, Step 2 fails if spot adjustments depend on w via the financing constraint $t^E(v, c) \leq w$, for then a non-stationary wage could create adjustment opportunities that a stationary wage fails to provide. A sufficient condition to rule out any interaction with the financing constraint is if the stationary w is large enough to compensate E for her benefit b for every possible inefficient order, as stated in (2). Although the parties may also wish to negotiate even if $\widetilde{m}^E$ is not inefficient but is less efficient than another action, the payments involved are smaller. That is, (2) is a sufficient condition because it states the largest payment that is required to compensate E, if for some realization of b and c all tasks happen to be inefficient and E and W would benefit from switching to no work. QED

Proof of Proposition 3: The expressions for ΔS and ΔV follow immediately from the discussion in the text. Result (1) follows because D^0 is increasing in β and b_H, D^E is decreasing in β and b_L, and U^0 is increasing in β, and all other derivatives with respect to β, b_H, and b_L are zero.

For result (2), note that $\frac{d\Delta W}{dc_H} + \frac{d\Delta W}{dc_L}$ and $\frac{d\Delta V}{dc_H} + \frac{d\Delta V}{dc_L}$ represent the effects of equal increases in c_H and c_L that increase $\overline{c}$ while leaving Δc unchanged. For ΔS, the result follows because D^0 is decreasing and D^E increasing in c_H. These are also the only terms affecting ΔV because U^0 and the obedience wage premium $(1-\delta)\gamma\Delta c$ depend only on Δc but not the level of costs. QED

Proof of Proposition 4: Part (a): $w_0(\delta)$ is decreasing in δ with $w_0(1) = \overline{c} - U^E$. Thus, as long as $b_L \leq \overline{c} - U^E$, spot adjustments are feasible for all δ and the optimal wage is $w_0(\delta)$. Parts (b) and (c): If $b_L > \overline{c} - U^E$, then $w_0(\widehat{\delta}) = b_L$ for some $\widehat{\delta} < 1$. For $\delta > \widehat{\delta}$, E can set $w = b_L$, enabling spot adjustments, or forgo adjustments and set $w = w_1(\delta)$. Since $V^E = b^E - w$ from (3), the former option is best if and only if $b_L \leq w_1(\delta) = \overline{c} + (1-\delta)\gamma\Delta c$. If $b_L < w_1(1) = \overline{c}$, then $w = b_L$ is optimal for all $\delta \in (\widehat{\delta}, 1]$ (part b). If $b_L > w_1(1)$, then there exists $\widetilde{\delta} < 1$ such that $b_L = w_1(\widetilde{\delta})$. In that case, $w = b_L$ is optimal for $\delta \in (\widehat{\delta}, \widetilde{\delta}]$, and $w = w_1(\delta)$ is optimal for $\delta \in (\widetilde{\delta}, 1]$ (part c). QED

Proof of Lemma 4: 1. Suppose $\Delta b \geq \Delta c$. (a) Prices: (i) E-type (b_L, b_L) can only trade with a worker with a low cost for at least one task, and therefore offers $t = (c_L, c_L)$. (ii) For type (b_H, b_L), candidates for prices are $t = (c_H, c_L)$ or $t = (c_L, c_L)$. If E offers (c_H, c_L), all types of W accept task A, and E's profit is $b_H - c_H$. If E offers (c_L, c_L), then (resolving ties in favor of A), types (c_L, c_L) and (c_L, c_H) accept A, type (c_H, c_L) accepts B and type (c_H, c_H) rejects both prices. E's resulting expected profit is $\gamma(b_H - c_L) + q_M(b_L - c_L)$. Then $t = (c_L, c_L)$ is optimal if

$$\gamma(b_H - c_L) + q_M(b_L - c_L) > b_H - c_H \quad \text{or} \quad \gamma > \widehat{\gamma} - \frac{b_L - c_L}{b_H - c_L} q_M = \gamma_1.$$

(iii) If E-Type (b_H, b_H) offers $t = (c_H, c_H)$, she will trade for sure and receive profit $b_H - c_H$. If she offers $t = (c_H, c_L)$, then all W-types except (c_H, c_L) except task A, whereas type (c_H, c_L) is indifferent and therefore picks the more efficient task B. E's resulting expected payoff is $(1 - q_M)(b_H - c_H) + q_M(b_H - c_L)$, which always exceeds the payoff from offering $t = (c_H, c_H)$. If E offers $t = (c_L, c_L)$, finally, types (c_L, c_L) and (c_L, c_H) accept task A, (c_H, c_L) accepts B, and type (c_H, c_H) rejects both, leading to a payoff for E of $(1 - q_H)(b_H - c_L)$. This option is preferred to $t = (c_H, c_L)$ if

$$\begin{aligned} (1 - q_H)(b_H - c_L) &> (1 - q_M)(b_H - c_H) + q_M(b_H - c_L) \\ \text{or} \quad \gamma(b_H - c_L) &> (1 - q_M)(b_H - c_H), \end{aligned} \tag{18}$$

which is equivalent to $\gamma > (1 - q_M)\widehat{\gamma} = \gamma_0$. It is straightforward to check that $\gamma_0 < \gamma_1$ if and only if $\Delta b > \Delta c$ as assumed, which leads to the three intervals stated in the proposition, and the optimal prices for each type of E.

(b) U^0 and D^0: (i) $\gamma \leq \gamma_0$: Type (b_L, b_L) offers $t = (c_L, c_L)$ and trades with W if and only if it is efficient, leading to $U^0 = D^0 = 0$. Type (b_H, b_L) offers $t = (c_H, c_L)$ and trades with all types of W, which is efficient $(D^0 = 0)$ and leaves a rent Δc to a W with low cost for task A (which occurs with probability γ). The same holds analogously for type (b_L, b_H). Type (b_H, b_H) offers $t = (c_H, c_L)$ and all W pick task A, except for type (c_H, c_L) who picks B. Again, the outcome is efficient and types (c_L, c_L) and (c_L, c_H) earn rent Δc. All in all, for $\gamma \leq \gamma_0$, $U^0 = (1 - p_L)\gamma\Delta c$ and $D^0 = 0$.

(ii) $\gamma \in [\gamma_0, \gamma_1)$: For types $(b_L, b_L), (b_H, b_L)$ and (b_L, b_H) the analysis is the same as in case (i). Type (b_H, b_H) now offers $t = (c_L, c_L)$ and all W with at least one low-cost task

pick that one, but earn no rent. Type (c_H, c_H) rejects both prices, which is inefficient. Collecting terms, we have $U^0 = 2p_M\gamma\Delta c$ and $D^0 = p_H q_H(b_H - c_H)$.

(iii) $\gamma \in [\gamma_1, 1)$: For types (b_L, b_L) and (b_H, b_H) the analysis is the same as in case (ii). Type (b_H, b_L) now offers $t = (c_L, c_L)$, which forces $U^0 = 0$. W-types (c_L, c_L) and (c_L, c_H) accept, which is efficient. Type (c_H, c_L) accepts task B, but task A is more efficient because $\Delta b > \Delta c$. Type (c_H, c_H) rejects, which is inefficient. The same holds analogously for E-type (b_L, b_H). Overall, we have $U^0 = 0$ and $D^0 = (1-p_L)q_H(b_H - c_H) + 2p_M q_M(\Delta b - \Delta c)$, where the first term captures inefficient failure to trade with (c_H, c_H) for all E-types except (b_L, b_L), and the second term captures the loss of surplus from ordering an inefficient task.

2. Now suppose $\Delta b < \Delta c$. (a) Prices: For E-types (b_L, b_L) and (b_H, b_H), the analysis is the same as in 1(a) above. If E-type (b_H, b_L) offers (c_H, c_L), all W-types but (c_H, c_L) accept task X, whereas type (c_H, c_L) is indifferent and chooses the more efficient task Y. E's profit is $(1 - q_M)(b_H - c_H) + q_M(b_L - c_L)$. If E offers (c_L, c_L), then W-types (c_L, c_L) and (c_L, c_H) accept X, type (c_H, c_L) accepts Y and type (c_H, c_H) rejects both prices. E's resulting expected profit is $\gamma(b_H - c_L) + q_M(b_L - c_L)$. Then $t = (c_L, c_L)$ is optimal if

$$\gamma(b_H - c_L) + q_M(b_L - c_L) > (1 - q_M)(b_H - c_H) + q_M(b_L - c_L) \Longleftrightarrow \gamma > \gamma_0.$$

Thus, in the case $\Delta b < \Delta c$ there are only two intervals for γ to distinguish.

(b) For $\gamma \leq \gamma_0$, the analysis of Lemma 4 applies without change: $U^0 = (1 - p_L)\gamma\Delta c$ and $D^0 = 0$. For $\gamma > \gamma_0$, the analysis is the same as in the case $\gamma > \gamma_1$ above, except for E-type (b_H, b_L) who offers $t = (c_L, c_L)$. As before, type (c_H, c_H) rejects, which is inefficient, but all other W-types, including (c_H, c_L), accept the efficient task. Consequently, we have $U^0 = 0$ as before, but the deadweight loss is only $D^0 = (1 - p_L)q_H(b_H - c_H)$. QED

Proof of Lemma 5: 1. $\Delta b \geq \Delta c$: Without loss of generality assume that E-types (b_L, b_L) and (b_H, b_H) order X unless Y is more efficient, i.e. if W has type (c_H, c_L). These orders are efficient except when $b = (b_L, b_L)$ and $c = (c_H, c_H)$, which occurs with probability $p_L q_H$. Types (b_H, b_L) and (b_L, b_H) order task X and Y, respectively, which is efficient for all types of W because $\Delta b \geq \Delta c$ (that is, with $b = (b_H, b_L)$ and $c = (c_H, c_L)$, for instance, X is efficient). W-type (c_H, c_H) can offer E $t = b_L$ in return for switching to no work, which E accepts if and only if $b = (b_L, b_L)$. Thus, the outcome is efficient in all states ($D^E = 0$), and W's expected adjustment gain is $U^E = p_L q_H(c_H - b_L)$.

It remains to determine W's expected cost. If $b = (b_H, b_L)$, all types of W execute X,

at expected cost $\overline{c}$. The same holds for $b = (b_L, b_H)$. If $b = (b_L, b_L)$ or (b_H, b_H), E orders A except when $c = (c_H, c_L)$. As a result, W's cost in these b-states is lower than $\overline{c}$ by $q_M \Delta c$. Overall, therefore, $c^E = \overline{c} - (p_L + p_H) q_M \Delta c$, which leads to $c_H - c^E = [\gamma + (p_L + p_H) q_M] \Delta c$.

2. $\Delta b < \Delta c$: For E-types (b_L, b_L) and (b_H, b_H), the above analysis applies without change. Type (b_H, b_L) again orders task X, but that is inefficient if $c = (c_H, c_L)$. But that W-type can infer from being ordered X that he is facing E-type (b_H, b_L), because every other type of E would have ordered Y for reasons of profitability or efficiency. Knowing which E-type he is facing, this W can offer $t = \Delta b$ to persuade E to switch to Y, which E accepts and which yields W a gain of $\Delta c - \Delta b$. We therefore have $D^E = 0$ and $U^E = p_L q_H (c_H - b_L) + 2 p_M q_M (\Delta c - \Delta b)$. W's expected cost *before* any spot adjustments is the same as above. QED

Proof of Proposition 5: $\frac{d\Delta V}{db_H} + \frac{d\Delta V}{db_L}$ represents the effects of equal increases in b_H and b_L that increase $\overline{b}$ while leaving Δb unchanged, and $\frac{d\Delta V}{dc_H} + \frac{d\Delta V}{dc_L}$ represents the effects of equal increases in c_H and c_L that increase $\overline{c}$ while leaving Δc unchanged. We can examine these changes separately for V^0 and V^E. By inspection of Table 1, $\frac{d(U^0 + D^0)}{db_H} + \frac{d(U^0 + D^0)}{db_L} \geq 0$ and $\frac{d(U^0 + D^0)}{dc_H} + \frac{d(U^0 + D^0)}{dc_L} \leq 0$ for all five cases shown, and therefore $\frac{dV^0}{db_H} + \frac{dV^0}{db_L} \leq 0$ and $\frac{dV^0}{dc_H} + \frac{dV^0}{dc_L} \geq 0$. In fact, all derivatives applied to U^0 alone are zero; the reported derivatives of $U^0 + D^0$ are those of D^0. With employment, we need inspect only the expression $x = c_H - c^E + U^E$ in $\widetilde{V}^E$, and inspection of the expressions given in Proposition 5 establishes $\frac{dx}{db_H} + \frac{dx}{db_L} \leq 0$ and $\frac{dx}{dc_H} + \frac{dx}{dc_L} \geq 0$ therefore $\frac{dV^E}{db_H} + \frac{dV^E}{db_L} \geq 0$ and $\frac{dV^E}{dc_H} + \frac{dV^E}{dc_L} \leq 0$. Putting the results for V^0 and V^E together establishes the result for ΔV.

As for ΔS, since employment is always efficient, we need consider only changes in the spot-trade deadweight loss D^0, which were already reported above. Therefore, $\frac{dS^0}{db_H} + \frac{dS^0}{db_L} \leq 0$ and $\frac{dS^0}{dc_H} + \frac{dS^0}{dc_L} \geq 0$, which establishes the result for ΔS. QED

Proof of Proposition 6: (a) Looking at a simultaneous change $\frac{\partial p_H}{\partial \varepsilon_b} = \frac{\partial p_L}{\partial \varepsilon_b} = 1$ and $\frac{\partial p_M}{\partial \varepsilon_b} = -1$, we have $\frac{\partial \Delta V}{\partial \varepsilon_b} = \frac{\partial \Delta V}{\partial p_H} + \frac{\partial \Delta V}{\partial p_L} - \frac{\partial \Delta V}{\partial p_M}$. Consider first spot trade with $V^0 = W^* - U^0 - D^0$. We can ignore W^* because it cancels out in ΔV; the expressions for U^0 and D^0 are stated in Lemma 4. For $\gamma \leq \gamma_0$, $\frac{\partial (U^0 + D^0)}{\partial \varepsilon_b} = -\gamma \Delta c < 0$. For $\gamma \in [\gamma_0, \gamma_1)$, $\frac{\partial (U^0 + D^0)}{\partial \varepsilon_b} = q_H (b_H - c_H) - 2\gamma \Delta c$. Because $\gamma > \gamma_0$, condition (18) holds, which using $b_H - c_L = b_H - c_H + \Delta c$ can be rewritten as $\gamma \Delta c > q_H (b_H - c_H)$. It follows that $\frac{\partial (U^0 + D^0)}{\partial \varepsilon_b} < 0$ for this range of γ as well. For $\gamma \in [\gamma_1, 1)$, finally, $\frac{\partial (U^0 + D^0)}{\partial \varepsilon_b} = -q_H (b_H - c_H) - 2 q_M (\Delta b - \Delta c) < 0$. Thus, for any γ, $\frac{\partial (U^0 + D^0)}{\partial \varepsilon_b} < 0$ and therefore $\frac{\partial V^0}{\partial \varepsilon_b} > 0$.

The employment profit V^E depends on $c_H - c^E + U^E$. If $\Delta b \geq \Delta c$, the derivative of this expression with respect to ε_b is $2q_M\Delta c + q_H(c_H - b_L) > 0$. If $\Delta b < \Delta c$, the derivative is $2q_M\Delta c + q_H(c_H - b_L) + 2q_M(\Delta c - \Delta b) > 0$. In both cases, therefore, $\frac{\partial V^E}{\partial \varepsilon_b} < 0$, and overall, it follows that $\frac{\partial \Delta V}{\partial \varepsilon_b} < 0$.

(b) Looking at a simultaneous change $\frac{\partial q_H}{\partial \varepsilon_c} = \frac{\partial q_L}{\partial \varepsilon_c} = 1$ and $\frac{\partial q_M}{\partial \varepsilon_c} = -1$, we have $\frac{\partial \Delta V}{\partial \varepsilon_c} = \frac{\partial \Delta V}{\partial q_H} + \frac{\partial \Delta V}{\partial q_L} - \frac{\partial \Delta V}{\partial q_M}$. Consider first spot trade. For $\gamma \leq \gamma_0$, $\frac{\partial (U^0 + D^0)}{\partial \varepsilon_b} = 0$. For $\gamma \in [\gamma_0, \gamma_1)$, $\frac{\partial (U^0 + D^0)}{\partial \varepsilon_b} = p_H(b_H - c_H) > 0$. For $\gamma \in [\gamma_1, 1)$,

$$\begin{aligned} \frac{\partial (U^0 + D^0)}{\partial \varepsilon_b} &= (1 - p_L)(b_H - c_H) - 2p_M(\Delta b - \Delta c) \\ &> 2p_M(b_H - c_H) - 2p_M(b_H - c_H - (b_L - c_L)) \\ &= 2p_M(b_L - c_L) > 0. \end{aligned}$$

Thus, for any γ, $\frac{\partial (U^0 + D^0)}{\partial \varepsilon_c} > 0$ and therefore $\frac{\partial V^0}{\partial \varepsilon_c} < 0$. For employment and if $\Delta b \geq \Delta c$, the derivative of $c_H - c^E + U^E$ with respect to ε_c is $-(p_L + p_H)\Delta c + p_L(c_H - b_L)$, which is negative because $b_L > c_L$ and therefore $\Delta c > c_H - b_L$. If $\Delta b < \Delta c$, the derivative is $-(p_L + p_H)\Delta c + p_L(c_H - b_L) - 2p_M(\Delta c - \Delta b) < 0$. In both cases, therefore, $\frac{\partial V^E}{\partial \varepsilon_c} > 0$ and overall, $\frac{\partial \Delta V}{\partial \varepsilon_c} > 0$. QED

Proof of Proposition 7: Preliminary analysis: The analysis of E-authority is covered by Lemma 5. If $\Delta b \geq \Delta c$, then all of E's orders are efficient except when $(b, c) = (b_L, b_L, c_H, c_H)$. If $\Delta b < \Delta c$, E's orders are additionally inefficient when $b = (b_H, b_L)$ and $c = (c_H, c_L)$ or vice versa. No matter the ranking of Δb and Δc, W's expected cost from E's default orders is $c^E = \overline{c} - (p_L + p_H)q_M\Delta c$.

If W has authority, then types (c_L, c_H) and (c_H, c_L) pick task X and Y, respectively, whereas types (c_L, c_L) and (c_H, c_H) pick whichever task has the higher benefit (resolving remaining ties in favor of X). Thus, W's cost is c_L if at least one cost is low, and c_H if $c = (c_H, c_H)$, which results in the expected cost (before adjustments) of $c^E = \overline{c} - q_M\Delta c$, which is smaller than with E-authority. If $b = (b_L, b_L)$ and $c = (c_H, c_H)$, neither task is efficient but W is required to provide his service, thus W's choice is inefficient. If $\Delta b \geq \Delta c$, two more inefficient choices occur if $b = (b_H, b_L)$ and $c = (c_H, c_L)$ or vice versa. $\Delta b < \Delta c$, then W's pick of the lower-cost task is efficient, and only $(b, c) = (b_L, b_L, c_H, c_H)$ leads to inefficiency.

Part (a): Without spot adjustments ($U^E = 0$) and if $\Delta b \geq \Delta c$, E-authority leads to

both a smaller deadweight loss and a smaller obedience wage premium $(1-\delta)(c_H - c^E)$ than W-authority, and is therefore more efficient and more profitable. If $\Delta b < \Delta c$, then W-authority leads to a smaller deadweight loss, but again a larger obedience wage premium. Specifically, with E-authority, E's profit is

$$\begin{aligned} & W^* - D^E - (1-\delta)(c_H - c^E) \\ = \; & W^* - p_L q_H (c_H - b_L) - 2p_M q_M (\Delta c - \Delta b) - (1-\delta)[\gamma + (p_L + p_H) q_M] \Delta c, \end{aligned}$$

whereas with W-authority, her profit is

$$W^* - p_L q_H (c_H - b_L) - (1-\delta)(\gamma + q_M)\Delta c.$$

Delegation is then optimal if

$$2(1-\delta)p_M q_M \Delta c < 2p_M q_M (\Delta c - \Delta b) \Longleftrightarrow (1-\delta)\Delta c < \Delta c - \Delta b \Longleftrightarrow \delta > \Delta b / \Delta c,$$

as stated.

Part (b): Suppose now that either side can suggest an adjustment as long as the price offer to the other party is nonnegative. If $\Delta b \geq \Delta c$, then the case of E-authority is covered by Lemma 5. If W has authority, three states are inefficient but who makes an offer depends on the state: If $c = (c_H, c_H)$, W can offer $t = b_L$ to switch to no work, which E accepts if and only if $b = (b_L, b_L)$. If $b = (b_H, b_L)$ and W picks Y, then E can offer $t = \Delta c$ to switch from Y to X, which only type $c = (c_H, c_L)$ will accept. The same holds analogously for the opposite case. As a result, we have $D^E = 0$ and $U^E = p_L q_H (c_H - b_L)$ and $V^E = p_M q_M (\Delta b - \Delta c)$, where the latter does not appear in (10) because it is absorbed in S^*. E-authority is unambiguously more profitable both because while W's adjustment gain is the same in both cases, the obedience wage premium is larger with W-authority because c^E is smaller.

If $\Delta b < \Delta c$, then the case of E-authority is again covered by Lemma 5 because adjustments away from all three inefficient orders involve payments from W to E. If W has authority, only the state $(b, c) = (b_L, b_L, c_H, c_H)$ is inefficient, in which case the parties switch to no work following payment of b_L by W to E. Thus, $D^E = 0$ and $U^E = p_L q_H (c_H - b_L)$. Given the general expression $\widetilde{V}^B = W^* - (1-\delta)(c_H - c^E + U^E)$, we need

only compare the $c_H - c^E + U^E$ term for each case, and E-authority is more profitable if

$$c_H - [\overline{c} - (p_L + p_H) q_M \Delta c] + p_L q_H (c_H - b_L) + 2 p_M q_M (\Delta c - \Delta b) < c_H - (\overline{c} - q_M \Delta c) + p_L q_H (c_H - b_L),$$

which upon simplification turns out to always hold. QED

Proof of Proposition 8: 1. Efficient tasks: Assumption (15) ensures that task Y is efficient if and only if $b^Y = b + b_H$. Specifically, for X to be efficient in the case $b^Y = b + b_L$ if even if $c^X = c_H$ and $c^Y = c_L + c$ requires $b - c_H > b + b_L - c_L - c$ or $c > b_L + \Delta c$. And for Y to be efficient in the case $b^Y = b + b_H$ even if $c^X = c_L$ and $c^Y = c_H + c$ requires $b + b_H - c - c_H > b - c_L$ or $c < b_H - \Delta c$.

2. Consider a contract with $A = \{X, Y\}$ and a uniform wage w. Then because $b^Y > b^X = b$, E will always order Y. Regardless of type, W can offer $t = b_L$ (which is $\min\{b^Y\} - b^X$) to switch to X, which E accepts iff $b^Y = b + b_L$, and which restores efficiency. W gains $c - b_L$ regardless of type, gains an additional Δc if $c = (c_L, c_H + c)$, but loses an additional Δc if $c = (c_H, c_L + c)$. The latter two effects cancel out, and we have $U^E = (1 - \beta)(c - b_L)$. W's expected cost from E's default order is $c^E = \overline{c} + c$ (the average cost of Y) and the highest possible cost is $c_H + c$. Therefore, $c_H - c^E = \gamma \Delta c$. E's resulting profit is

$$\widetilde{V}^E = W^* - (1 - \delta)[\gamma \Delta c + (1 - \beta)(c - b_L)]. \tag{19}$$

3. Now suppose $A = \{X\}$;X is the default order and therefore $c^E = \overline{c}$, but E and W can renegotiate to Y. W would never pay to switch to Y, but E would pay W; however, E does not know W's cost. If E offers $t = c + \Delta c$, all types of W, even $c = (c_L, c_H + c)$ accept, and E's adjustment gain is $b_H - c - \Delta c$. If $t = c$, then all W-types except for $c = (c_L, c_H + c)$ accept, and E's adjustment gain is $(1 - q_M)(b_H - c)$. If $t = c - \Delta c$, only type $c = (c_H, c_L - c)$ accepts, and E's adjustment gain is $q_M(b_H - c + \Delta c)$. Then $t = c$ dominates $t = c - \Delta c$ if

$$(1 - q_M)(b_H - c) \geq q_M(b_H - c + \Delta c) \Longleftrightarrow (1 - 2q_M)(b_H - c) \geq q_M \Delta c.$$

Since $b_H - c > \Delta c$ per (15), a sufficient condition is $1 - 2q_M \geq q_M$ or $q_M \leq 1/3$. If c^X and c^Y are independent, $q_M = \gamma(1 - \gamma) \leq 1/4$. Therefore, a nonnegative correlation of costs,

$q_M \leq \gamma(1-\gamma)$, is sufficient for $q_M \leq 1/3$ to hold. And $t = c$ dominates $t = c + \Delta c$ if

$$(1 - q_M)(b_H - c) \geq b_H - c - \Delta \Longleftrightarrow q_M < \frac{\Delta c}{b_H - c}.$$

With a nonnegative correlation of costs and hence $q_M \leq 1/4$, this condition is satisfied if $4\Delta c \geq b_H - c$, the condition stated in the proposition. Thus, under the assumptions stated, for E with $b^Y = b + b_H$ wishing to switch to Y, offering $t = c$ is optimal. In this case, only W-type $(c_H, c_L + c)$ earns a rent, and therefore $U^E = \beta q_M \Delta c$. E's profit is

$$\widetilde{V}^E = W^* - (1 - \delta)[\gamma \Delta c + \beta q_M \Delta c]. \tag{20}$$

Comparing (20) with (19), a contract with $A = \{X\}$ is more profitable if $\beta q_M \Delta c < (1 - \beta)(c - b_L)$.

4. Now suppose $A = \{X, Y\}$ but the wage can be tied to E's order. Setting $w^X = w$ and $w^Y = w + c$ accomplishes two things: first, if the obedience constraint (4) holds with equality for task X at wage w, then it also holds with equality at wage $w + c$ for task Y. Thus, there is no unnecessary slack in the incentive constraints. Second, with E choosing orders to maximize $b^k - w^k$, E will order Y if and only if $b^Y = b + b_H$, as is efficient. This follows from $b_H > c > b_L$ as implied by (15). Because E's orders are efficient, $U^E = 0$, and we have $\widetilde{V}^E = W^* - (1 - \delta)\gamma \Delta c$, which exceeds the profit of both other contractual options. QED

Proof of Proposition 9: 1. Suppose $\Delta b \geq \Delta c$. 1a. Consider $A = \{X, Y\}$. E's orders are the same as in the symmetric case covered by Lemma 5. In particular, all orders except when $b = (b_L, b_L)$ and $c = (c_H, c_H)$ are efficient, and W can in the latter case negotiate to no work and earn an expected adjustment gain of $U^E = p_{LL} q_{HH} (c_H - b_L)$. To compute W's expected cost c^E, note that E will—as a result of picking a profit-maximizing task that is Pareto-efficient—order a task with the lowest possible cost, except when $(b, c) = (b_H, b_L, c_H, c_L)$ or vice versa. The expected value of the lowest cost is $c_{\min} = (1 - q_{HH})c_L + q_{HH} c_H$, and then W's expected cost from default orders is $c^E = c_{\min} + (p_{LH} q_{LH} + p_{HL} q_{HL})\Delta c$, which leads to

$$\begin{aligned} c_H - c^E &= (1 - q_{HH} - p_{LH} q_{LH} - p_{HL} q_{HL})\Delta c. \\ &= [q_{LL} + q_{LH} + (1 - p_{HL}) q_{HL} - p_{LH} q_{LH}]\Delta c \end{aligned} \tag{21}$$

1b. If instead $A = \{X\}$, E orders X, and W's expected cost is $c^E = (q_{LL} + q_{LH})c_L + (q_{HL} + q_{HH})c_H$, which leads to

$$c_H - c^E = (q_{LL} + q_{LH})\Delta c.$$

X is inefficient in 7 of 16 states: (i) as usual, if $(b, c) = (b_L, b_L, c_H, c_H)$; (ii) 4 states in which $b = (b_L, b_H)$; and (iii) 3 states in which $c = (c_H, c_L)$—not including when $b = (b_H, b_L)$ and therefore X is efficient—, one of which overlaps with the previous 4, namely $(b, c) = (b_L, b_H, c_H, c_L)$.

Restricting bargaining to nonnegative offers by either side, spot bargaining restores efficiency in all states but one: (i) W-type $c = (c_H, c_H)$ can offer $t = b_L$ to switch to no work, which E accepts iff $b = (b_L, b_L)$. (ii) W-type $c = (c_H, c_L)$ can offer $t = 0$, which E accepts if $b \in \{(b_L, b_L), (b_H, b_H), (b_L, b_H)\}$, i.e. when it is efficient, whereas $b = (b_H, b_L)$ efficiently rejects (and there is no reason for W to pay more).

(iii) E-type $b = (b_L, b_H)$ has two relevant options to persuade W to switch to Y: if $t = 0$, all W-types except $c = (c_L, c_H)$ will accept, giving E an adjustment gain of $(1 - q_{LH})\Delta b$. If $t = \Delta c$, then all W-types (as would be efficient) accept, and E's gain is $\Delta b - \Delta c$. Offering $t = 0$ is more profitable if $(1 - q_{LH})\Delta b \geq \Delta b - \Delta c$ or $\Delta c/\Delta b \geq q_{LH}$, which is equivalent to the condition $\Delta b/\Delta c \leq 1/q_{LH}$ stated in the proposition. In the state $(b, c) = (b_L, b_H, c_H, c_L)$, both E and W offer $t = 0$ to the other and thus agree to switch to Y. In the state $(b, c) = (b_L, b_H, c_L, c_H)$, however, switching to Y is efficient but does not occur because E optimally offers $t = 0$, which W rejects. The resulting expected deadweight loss is $D^E = p_{LH}q_{LH}(\Delta b - \Delta c)$. Collecting terms, W gains $c_H - b_L$ if $c = (c_H, c_H)$ and gains $(1 - p_{HL})\Delta c$ if $c = (c_H, c_L)$, leading to $U^E = p_{LL}q_{HH}(c_H - b_L) + (1 - p_{HL})q_{HL}\Delta c$.

1c. From the preceding analysis, E's profit if $A = \{X, Y\}$ is

$$\begin{aligned} \widetilde{V}^E &= W^* - (1 - \delta)[c_H - c^E + U^E] \\ &= W^* - (1 - \delta)\{[q_{LL} + q_{LH} + (1 - p_{HL})q_{HL} - p_{LH}q_{LH}]\Delta c + p_{LL}q_{HH}(c_H - b_L)\}, \end{aligned} \tag{22}$$

whereas her profit with $A = \{X\}$ is

$$\begin{aligned}\widetilde{V}^E &= W^* - D^E - (1-\delta)[c_H - c^E + U^E] \\ &= W^* - p_{LH}q_{LH}(\Delta b - \Delta c) \\ &\quad -(1-\delta)[(q_{LL}+q_{LH})\Delta c + p_{LL}q_{HH}(c_H - b_L) + (1-p_{HL})q_{HL}\Delta c]. \qquad (23)\end{aligned}$$

Comparing (23) with (22), a contract with $A = \{X, Y\}$ is unambiguously more profitable.

2. Suppose $\Delta b < \Delta c$. 2a. Consider $A = \{X, Y\}$. E's orders are the same as in the symmetric case covered by Lemma 5. In all three states with inefficient orders, W can negotiate to the efficient action and earn an expected adjustment gain of $U^E = p_{LL}q_{HH}(c_H - b_L) + (p_{LH}q_{LH} + p_{HL}q_{HL})(\Delta c - \Delta b)$. W's expected cost c^E is the same as in the case 1a above, and (21) holds.

2b. If instead $A = \{X\}$, E orders X, and as before we have $c_H - c^E = (q_{LL}+q_{LH})\Delta c$. X is inefficient in 7 of 16 states: (i) $(b, c) = (b_L, b_L, c_H, c_H)$; (ii) 3 states in which $b = (b_L, b_H)$, except for when $c = (c_L, c_H)$; and (iii) 4 states in which $c = (c_H, c_L)$, one of which overlaps with the previous 3, namely $(b, c) = (b_L, b_H, c_H, c_L)$. Spot bargaining with nonnegative offers by either side again restores efficiency in all states but one: (i) W-type $c = (c_H, c_H)$ offers $t = b_L$ to switch to no work, which E accepts iff $b = (b_L, b_L)$. (ii) E-type $b = (b_L, b_H)$ can offer $t = 0$ to switch to Y, which W accepts if $c \in \{(c_L, c_L), (c_H, c_H), (c_H, c_L)\}$, i.e. when it is efficient, whereas $c = (c_L, c_H)$ efficiently rejects (and there is no reason for E to pay more).

(iii) W-type $c = (c_H, c_L)$ has two relevant options to persuade E to switch to Y: if $t = 0$, all E-types except $b = (b_H, b_L)$ will accept, giving W an adjustment gain of $(1-p_{HL})\Delta c$. If $t = \Delta b$, then all E-types (as would be efficient) accept, and W's gain is $\Delta c - \Delta b$. Offering $t = 0$ is more profitable if $(1-p_{HL})\Delta c \geq \Delta c - \Delta b$ or $\Delta b/\Delta c \geq p_{HL}$, the condition stated in the proposition. In the state $(b, c) = (b_L, b_H, c_H, c_L)$, both E and W offer $t = 0$ to the other and thus agree to switch to Y. In the state $(b, c) = (b_H, b_L, c_H, c_L)$, however, switching to Y is efficient but does not occur because W optimally offers $t = 0$, which E rejects. The resulting expected deadweight loss is $D^E = p_{HL}q_{HL}(\Delta c - \Delta b)$. Collecting terms, we have $U^E = p_{LL}q_{HH}(c_H - b_L) + (1-p_{HL})q_{HL}\Delta c$.

2c. Except for the deadweight loss with $A = \{X\}$ occuring in a different state than when $\Delta b \geq \Delta c$, the comparison of E's profit if $A = \{X, Y\}$ and if $A = \{X\}$ is the same as in step 1c, and leads to the conclusion that $A = \{X, Y\}$ is more profitable.

Proof of Proposition 10: If $b = b_L$, plausible price offers to the two workers are $t = (c_L, c_L)$ and $t = (c_H, c_L)$. (With $t = (c_H, c_H)$, both workers would accept but the entrepreneur would make a loss of $2b_L - 2c_H$.) If $t = (c_L, c_L)$, then both workers—as required for positive output—will accept only if $c_1 = c_2 = c_L$, leading to a profit for E of $2\gamma^2(b_L - c_L)$ as opposed to $2\gamma(b_L - c_L)$ with independent production. Thus, the average price is the same as with independent production, but the profit strictly lower.

If $b = b_L$ and E offers $t = (c_H, c_L)$, then W1 accepts for sure and W2 iff $c_2 = c_L$, leading to profit $\gamma(2b_L - c_H - c_L)$ for E, which is less than $2\gamma(b_L - c_L)$ with independent production. Though this strategy can be optimal for some parameter values, the average price is strictly higher and the profit lower than with independent production.

If $b = b_H$, all three different price pairs are plausible candidates for E to offer. If E offers $t = (c_L, c_L)$, both workers accept iff $c_1 = c_2 = c_L$, and E's profit is $2\gamma^2(b_H - c_L)$. If E offers $t = (c_H, c_L)$, then W1 accepts for sure and W2 iff $c_2 = c_L$, leading to profit $\gamma(2b_H - c_H - c_L)$ for E. This option is more profitable than $t = (c_L, c_L)$ if $\gamma(2b_H - c_H - c_L) \geq 2\gamma^2(b_H - c_L)$ or

$$\gamma \leq \frac{2b_H - c_H - c_L}{2(b_H - c_L)} = \widehat{\gamma}_2.$$

Finally, if $t = (c_H, c_H)$, both workers accept for sure and E's profit is $2(b_H - c_H)$, which is no smaller than $\gamma(2b_H - c_H - c_L)$ iff

$$\gamma \leq \frac{2(b_H - c_H)}{2b_H - c_H - c_L} = \widehat{\gamma}_1.$$

For $\widehat{\gamma}$ as defined in Lemma 1, it is straightforward to show that $\widehat{\gamma} < \widehat{\gamma}_1 < \widehat{\gamma}_2$, which leads to four intervals for γ: (i) If $\gamma \leq \widehat{\gamma}$, then $t = (c_H, c_H)$ with both independent and complementary production, and E's profit in both cases is $2(b_H - c_H)$. (ii) If $\gamma > \widehat{\gamma}$, then with independent production, $t = (c_L, c_L)$ and E's profit is $2\gamma(b_H - c_L)$. Now if $\gamma \in (\widehat{\gamma}, \widehat{\gamma}_1]$, then with complementary production, $t = (c_H, c_H)$ and the resulting profit is $2(b_H - c_H)$, which is less than $2\gamma(b_H - c_L)$ because $\gamma > \widehat{\gamma}$. (iii) If $\gamma \in (\widehat{\gamma}_1, \widehat{\gamma}_2]$, then with complementary production, $t = (c_H, c_L)$ and the resulting profit is $\gamma(2b_H - c_H - c_L) < 2\gamma(b_H - c_L)$. (iv) If $\gamma > \widehat{\gamma}_2$, then with complementary production, $t = (c_L, c_L)$ and the resulting profit is $2\gamma^2(b_H - c_L) < 2\gamma(b_H - c_L)$. Thus, in case (i), prices and profits are the same with independent and complelementary production, whereas in all other cases, the average price is higher and the profit lower with complementary production. QED

Proof of Proposition 11: With employment, both workers are required to work. If $b = b_H$, that is efficient for all worker types. If $b = b_L$, it's efficient if $c_1 = c_2 = c_L$ and inefficient if $c_1 = c_2 = c_H$. If $c_1 = c_H$ and $c_2 = c_L$ or vice versa, for both workers to work is efficient iff $2b_L \geq c_H + c_L$.

A high-cost worker can offer $t = b_L$ for E to switch to no work, and if both workers do so and $b = b_L$, then E will accept, which restores efficiency in that state. However, if $2b_L < c_H + c_L$, then in the state $(b, c_1, c_2) = (b_L, c_H, c_L)$ (or vice versa), only one worker will offer $t = b_L$, and E won't accept. Restoring efficiency in that case would require $2b_L \leq c_H$, in which case one worker's offer of $t = 2b_L$ would suffice for E to accept to switch to no work. Suppose $2b_L < b_H$, in which case only $b = b_L$ would ever accept $t = 2b_L$ (i.e., $b = b_H$ would reject even if $t = 2b_L$ was offered by both workers). Then, focusing on the case $b = b_L$, offering $t = 2b_L$ would lead to an adjustment gain of $c_H - 2b_L$ irrespective of the other worker's type, whereas offering $t = b_L$ would lead to an adjustment gain of $(1-\gamma)(c_H - b_L)$ corresponding to the case in which agreement occurs only if the other worker has a high cost as well. It follows that $t = 2b_L$ is optimal if $\gamma \geq b_L/(c_H - b_L)$, which is possible because $c_H \geq 2b_L$.

With independent production, E's profit per worker is $\widetilde{V}^E = S^* - (1-\delta)[\gamma\Delta c + (1-\beta)(1-\gamma)(c_H - b_L)]$ per Lemma 3. With complementary production and if $2b_L \geq c_H + c_L$, it is $\widetilde{V}^E = W^* - (1-\delta)[\gamma\Delta c + (1-\beta)(1-\gamma)^2(c_H - b_L)]$, which is higher because work is inefficient (and avoided) only in the state $(b, c_1, c_2) = (b_L, c_H, c_H)$. However, if $2b_L < c_H + c_L$ and the workers fail to negotiate away from work in the states (b_L, c_H, c_L) and (b_L, c_L, c_H), then E's profit per worker is

$$\widetilde{V}^E = W^* - (1-\beta)\gamma(1-\gamma)(c_H + c_L - 2b_L) - (1-\delta)[\gamma\Delta c + (1-\beta)(1-\gamma)^2(c_H - b_L)],$$

which because of the remaining deadweight loss is lower than with independent production. QED

Proof of Proposition 12: With one task, E's profit $\widetilde{V}^E$ from hiring one worker is given by Lemma 3, with

$$W^* = \beta(b_H - \overline{c}) + (1-\beta)\gamma(b_L - c_L). \tag{24}$$

With spot trade and E facing n workers competing, type b_L always offers $t = c_L$. Type

b_H can offer $t = c_H$ and make profit $b_H - c_H$, or offer $t = c_L$, which is accepted by at least one seller with probability $1 - (1-\gamma)^n$. If follows that $t = c_L$ is optimal if

$$1 - (1-\gamma)^n > \widehat{\gamma} \text{ or } \gamma \geq 1 - \sqrt[n]{1-\widehat{\gamma}},$$

where the last expression converges to 0 as $n \to \infty$. It follows that for any γ, there exists n for which E offers $t = c_L$ irrespective of her type. That offer is accepted with probability $1 - (1-\gamma)^n$, which converges to 1 for large n. Thus, E's profit from spot trade converges to $\overline{b} - c_L$, which exceeds (24) and therefore also $\widetilde{V}^E$. QED

References

Abraham, Katharine G., & Susan K. Taylor (1996): Firms' use of outside contractors: theory and evidence. *Journal of Labor Economics*, 14(3), 394-424.

Alchian, Armen, and Harold Demsetz (1972): Production, Information Costs, and Economic Organization. *American Economic Review* 62(5), 777-795.

Allen, George C. (1929). *The industrial development of Birmingham and the Black Country*, 1860-1927 (Vol. 10). London

Anderson, Tracy, and Peter Cappelli (2020): Management Without Managers: Expanding Our Understanding of Managing Through the Study of Contractors. Available at SSRN 3875634.

Asghar, Rob (2013): The Toughest Leadership Job Of All (It's Not What You Think), Forbes, November 15, 2013

Ausubel, Lawrence, Peter Cramton, and Raymond Deneckere (2002): Bargaining with incomplete information. In: Robert J. Aumann and Sergiu Hart, eds., *Handbook of Game Theory*, Vol. 3, Chapter 50, pp. 1897-1945. Amsterdam: Elsevier

Bajari, Patrick, Steven Tadelis (2001): Incentives versus transaction costs: A theory of procurement contracts. *Rand Journal of Economics*, 387-407.

Baker, George, Robert Gibbons, and Kevin J. Murphy (2002): Relational Contracts and the Theory of the Firm. *Quarterly Journal of Economics* 117 (1), pp. 39–84,

——— (2011): Relational adaptation. Working paper.

Baliga, Sandeep, and Tomas Sjöström (2018): A Theory of the Firm based on Haggling, Coordination and Rent-Seeking. Working paper.

Barnard, Chester (1938): *The Functions of the Executive*, Harvard University Press

Baron, James, and David Kreps (2013): Employment as an Economic and a Social Relationship. In: Gibbons and Roberts, pp. 315-341.

Barrero, Jose Maria, Nicholas Bloom, and Steven J. Davis (2021): Why working from home will stick. National Bureau of Economic Research, No. w28731

Barron, Dan, Jin Li, and Michal Zator (2019): Morale and Debt Dynamics. Working paper.

Bellos, David (2017). *The Novel of the Century: The Extraordinary Adventure of Les Misérables*. Farrar, Straus and Giroux.

Bertrand, Marianne, Chang-Tai Hsieh, and Nick Tsivanidis (2021): Contract labor and firm growth in India. National Bureau of Economic Research, No. w29151

Blau, Peter M. (1964): *Exchange and Power in Social Life*. Transaction Publishers

Bolton, Patrick, and Mathias Dewatripont (2005): *Contract Theory*, MIT Press

——— (2013): Authority in Organizations. In: Gibbons and Roberts, pp. 342-372.

Bolton, Patrick, and Ashvin Rajan (2003): The Employment Relation and the Theory of the Firm: Arm's Length Contracting vs Authority. Princeton University, working paper.

Braverman, Harry (1974): *Labor and Monopoly Capital: the Degradation of Work in the Twentieth Century*. New York, Monthly Review Press

Broughton, Kristin (2021): Demand for Part-Time CFOs Heats Up as Startups Raise More Money. *Wall Street Journal*, July 26, 2021

Bunch, Will (2018): In 1970s, workers at this GM plant tried to reinvent the American Dream. Instead, they watched it fade away. *Philadelphia Inquirer*, November 29, 2018

Buttrick, John (1952): The Inside Contract System. *Journal of Economic History*, 12(3), 205-221.

Cassar, Lea, and Stephan Meier (2018): Nonmonetary Incentives and the Implications of Work as a Source of Meaning. *Journal of Economic Perspectives*, 32(3), 215-38.

Chakravarty, Surejeet, and W. Bentley MacLeod, W. B. (2009): Contracting in the Shadow of the Law. *The RAND Journal of Economics*, 40(3), 533-557.

Chen, M. Keith, Peter Rossi, Judith Chevalier, and Emily Oehlsen (2019): The Value of Flexible Work: Evidence from Uber Drivers. *Journal of Political Economy* 127(6), pp. 2735-2794.

Cheung, Steven N. S. (1983): The contractual nature of the firm. *The Journal of Law and Economics* 26 (1), 1-21.

Chapman, Stanley D. (1974). The textile factory before Arkwright: A typology of factory development. *Business History Review*, 48(4), 451-478.

Clark, Gregory (1994). Factory discipline. *The Journal of Economic History*, 54(1), 128-163.

Coase, Ronald (1937): The Nature of the Firm. *Economica* 4 (16), pp. 386-405

——— (1960): The Problem of Social Cost. *Journal of Law and Economics* 3, pp. 1-44

——— (1992): The Institutional Structure of Production. *The American Economic Review*, 82(4), 713-719.

Coleman, James S. (1990): *Foundations of Social Theory*, Harvard University Press

Cunningham-Parmeter, Keith (2018): Gig-dependence: Finding the real independent contractors of platform work. *Northern Illinois University Law Review* 39: 379.

Department of Industrial Relations (2019): Independent contractor versus employee. State of California Department of Industrial Relations, retrieved from https://www.dir.ca.gov/dlse/f on Oct. 14, 2020

Dow, Gregory K. (1987): The function of authority in transaction cost economics. *Journal of Economic Behavior & Organization*, 8(1), 13-38.

Fama, Eugene. F. (1991): Time, salary, and incentive payoffs in labor contracts. *Journal of Labor Economics*, 9(1), 25-44.

Farrell, Joseph (1987): Information and the Coase theorem. *Journal of Economic Perspectives*, 1(2), 113-129.

Finkin, Matthew W. (2016): Beclouded work, beclouded workers in historical perspective. *Comparative Labor Law & Policy Journal*, 37(3), 603-618.

Ford Man, Sept. 20, 1917, p. 2-3 (Ford Motor Company).

Foss, Nicolai J. (2002): 'Coase vs Hayek': Economic Organization and the Knowledge Economy. International Journal of the Economics of Business, 9(1), 9-35.

Freeland, Robert F., & Zuckerman Sivan, Ezra W. (2018): The problems and promise of hierarchy: Voice rights and the firm. *Sociological Science*, 5, 143-181.

Gartenberg, Claudine M., and Todd Zenger (2021). The firm as a subsociety. Available at SSRN.

Gibbons, Robert (2005): Four formal(izable) theories of the firm? *Journal of Economic Behavior & Organization* 58 (2), pp. 200-245.

——— (2010): Transaction-cost economics: past, present, and future? *Scandinavian Journal of Economics* 112, no. 2, pp. 263-288.

——— and John Roberts (eds.) (2013): *Handbook of Organizational Economics*, Princeton University Press

Grossman, Sanford J., and Oliver D. Hart (1986): The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration. *Journal of Political Pconomy* 94 (4), pp. 691-719.

Hart, Oliver (1995): *Firms, Contracts, and Financial Structure.* Clarendon Press.

——— (1996): An economist's view of authority. *Rationality and Society*, 8(4), 371-386.

——— and John Moore (2008): Contracts as Reference Points. *Quarterly Journal of Economics* 123 (1). pp. 1-48.

Hay, Douglas, and Paul Craven (1993): Master and Servant in England and the Empire: A Comparative Study. *Labour/Le Travail* 31, pp. 175-184.

Holmström, Bengt (1999): The firm as a subeconomy. *Journal of Law, Economics, and Organization*, 15(1), 74-102.

———, and Paul Milgrom (1991): Multitask Principal-Agent Analyses: Incentive contracts, Asset Ownership, and Job Design. *Journal of Law and Economics* 7, pp. 24-52.

——— (1994). The firm as an incentive system. *American Economic Review*, 84 (4), pp. 972-991.

Holmström, Bengt, and John Roberts (1998). The boundaries of the firm revisited. *Journal of Economic Perspectives*, 12(4), 73-94.

Hu, S. Jack (2013): Evolving paradigms of manufacturing: from mass production to mass customization and personalization. *Procedia CIRP* 7, pp. 3-8.

Hudson, Pat (1981): Proto-industrialisation: the case of the West Riding Wool Textile Industry in the 18th and early 19th centuries. In: *History Workshop Journal* (Vol. 12, No. 1, pp. 34-61). Oxford University Press.

Internal Revenue Service (2017): Understanding employee vs. contractor designation. Retrieved from https://www.irs.gov/newsroom/understanding-employee-vs-contractor-designation on Oct. 14, 2020

Januszewski Forbes, Silke, and Mara Lederman (2009): Adaptation and vertical integration in the airline industry. *American Economic Review* 99(5), pp. 1831-49.

——— (2010): Does vertical integration affect firm performance? Evidence from the airline industry. *RAND Journal of Economics* 41 (4), pp. 765-790.

Juhász, Réka, Mara Squicciarini, and Nico Voigtländer (2020a): Technology Adoption and Productivity Growth: Evidence from Industrialization in France. National Bureau of Economic Research No. w27503

——— (2020b). Away from Home and Back: Coordinating (remote) Workers in 1800 and 2020. National Bureau of Economic Research No. w28251

Katz, Lawrence F., and Alan B. Krueger (2019): The rise and nature of alternative work arrangements in the United States, 1995–2015. *ILR Review* 72.2, pp. 382-416.

Kieser, Alfred (1994): Why organization theory needs historical analyses—and how this should be performed. *Organization Science*, 5(4), 608-620.

Klein, Benjamin, and Keith B. Leffler (1981): The role of market forces in assuring contractual performance. *Journal of Political Economy* 89 (4), 615-641.

Kranzberg, Melvin and Michael Hannan (2017): "History of the organization of work." *Encyclopedia Britannica*, 2 Jun. 2017, https://www.britannica.com/topic/history-of-work-organization-648000. Accessed 8 July 2021.

Landes, David S., ed. (1966): *The rise of capitalism.* Vol. 36742. Macmillan.

——— (1986): What Do Bosses Really Do? *Journal of Economic History*, 46, 585-625.

Lazear, Edward P., and Paul Oyer (2013). Personnel Economics. In: Gibbons and Roberts, pp. 479-519.

Lazerson, Mark (1995): A new phoenix?: Modern putting-out in the Modena knitwear industry. *Administrative Science Quarterly*, 34-59.

Levin, Jonathan (2003): Relational Incentive Contracts. *American Economic Review* 93 (3), pp. 835-857.

———, and Steven Tadelis (2010). Contracting for government services: Theory and evidence from US cities. *The Journal of Industrial Economics*, 58(3), 507-541.

Lewchuk, Wayne (1993). Men and monotony: Fraternalism as a managerial strategy at the Ford Motor Company. *The Journal of Economic History* 53(4), 824-856.

Li, Jin, and Niko Matouschek (2013): Managing conflicts in relational contracts. *American Economic Review* 103 (6), 2328-51.

Li, Jin, Arijit Mukherjee, and Luis Vasconcelos (2020): Rules in Relational Contracts. Working paper.

Lindbeck, Assar, and Dennis J. Snower (1996). Reorganization of firms and labor-market inequality. *The American Economic Review*, 86(2), 315-321.

——— (2000): Multitask learning and the reorganization of work: From tayloristic to holistic organization. *Journal of labor economics* 18 (3), pp. 353-376.

Malcomson, James (1999): Individual employment contracts. In: Orley Ashenfelter and David Card, *Handbook of Labor Economics*, Vol. 3B, Elsevier, pp. 2291-2372

Mantoux, Paul (1928): *The Industrial Revolution in the Eighteenth Century.* Harper Torchbooks, revised edition 1961

Marglin, Stephen (1974): What do Bosses do? The Origins and Functions of Hierarchy in Capitalist Production. *Review of Radical Political Economics* 6(2), 60-112.

Marino, Anthony M., John G. Matsusaka, and Ján Zábojník (2010): Disobedience and Authority. *Journal of Law, Economics, and Organization* 26 (3), pp. 427-459.

Marx, Karl (1867): *Capital. A Critique of Political Economy.* Retrieved from https://www.marxists.org/ Volume-I.pdf on Oct. 14, 2020

Masten, Scott E. (1986): Institutional choice and the organization of production: the make-or-buy decision. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, 493-509.

——— (1988): A legal basis for the firm. *Journal of Law, Economics, & Organization* 4.1 1988, pp. 181-198.

Matouschek, Niko (2004): Ex Post Inefficiencies in a Property Rights Theory of the Firm. *Journal of Law, Economics, and Organization* 20 (1), pp. 125–147.

McGregor, Douglas M. (1957). The human side of enterprise: American management association. *Management Review*, 1, 41-57.

McKelvey, Richard D., and Talbot Page (2002): Status Quo Bias in Bargaining: An Extension of the Myerson–Satterthwaite Theorem with an Application to the Coase Theorem. *Journal of Economic Theory* 107 (2), pp. 336-355.

McKendrick, Neil (1961): Josiah Wedgwood and factory discipline. *The Historical Journal* 4.1, pp. 30-55.

Milgrom, Paul, and John Roberts (1988): An economic approach to influence activities in organizations. *American Journal of Sociology*, 94, S154-S179.

——— (1990): Bargaining costs, influence costs, and the organization of economic activity. *Perspectives on Positive Political Economy*, 57, 60.

——— (1992): *Economics, Organization and Management.* Englewood Cliffs, NJ: Prentice-Hall.

Millward, Robert (1981): The emergence of wage labor in early modern England. *Explorations in Economic History* 18 (1), pp. 21-39.

Myerson, Roger, and Mark Satterthwaite (1983): Efficient Mechanisms for Bilateral Trade. *Journal of Economic Theory* 29, pp. 265-281.

Naidu, Suresh, and Noam Yuchtman, N. (2013): Coercive contract enforcement: law and the labor market in nineteenth century industrial Britain. *American Economic Review*, 103(1), 107-44.

National Labor Relations Board (2019): Advice Memorandum, Subject: Uber Technologies, Inc., Cases 13-CA-163062 et al. Retrieved from https://src.bna.com/Ibt on Oct. 14, 2020

Oyer, Paul (2016): The Independent Workforce in America. Study commissioned by Upwork. https://s3-us-west-1. amazonaws. com/adquiro-content-prod/documents/paul_oyer_the_in pdf.

Parsons, Donald O. (1986). The employment relationship: Job attachment, work effort, and the nature of contracts. In: *Handbook of Labor Economics*, 2, 789-848.

Paychex (2016): Goodbye, 9-5! The Growth of the Freelance Economy. Retrieved from https://www.paychex.com/articles/human-resources/goodbye-9-5-growth-of-the-freelance-economy, accessed on 7/21/21

Pollard, Sidney (1963): Factory discipline in the industrial revolution. *Economic History Review* 16 (2), pp. 254-271.

Posner, Eric A. (2020): The economic basis of the independent contractor/employee distinction. Working paper, University of Chicago.

Powell, Michael (2015): An influence-cost model of organizational practices and firm boundaries. *The Journal of Law, Economics, and Organization*, 31 (S1), i104-i142.

Rantakari, Heikki (2021): Simon Says? Equilibrium Obedience and the Limits of Authority. Working paper, University of Rochester.

Rebitzer, James B. (1993): Radical political economy and the economics of labor markets. *Journal of Economic Literature* 31 (3), 1394-1434.

Roberts, John (2004). *The modern firm: Organizational Design for Performance and Growth.* Oxford University Press.

Segal, Ilya, and Michael D. Whinston. (2013): Property rights. In: Gibbons and Roberts, pp. 100-158

——— (2016): Property Rights and the Efficiency of Bargaining. *Journal of the European Economic Association* 14 (6), pp. 1287–1328

Siddiqui, Faiz (2020): Uber and Lyft must make their drivers in California full employees, judge rules. *Washington Post*, August 10, 2020

Simon, Herbert (1945): *Administrative Behavior.* New York (Free Press)

——— (1951): A Formal Theory of the Employment Relationship. *Econometrica* 19 (3), pp. 293-305.

——— (1991): Organizations and markets. *Journal of Economic Perspectives* 5 (2), 25-44.

Stigler, George J. (1951). The Division of Labor is Limited by the Extent of the Market. *Journal of Political Economy*, 59(3), 185-193.

Stone, Dan, Edward Deci, and Richard Ryan (2009): Beyond talk: Creating autonomous motivation through self-determination theory. *Journal of General Management*, 34(3), 75-91.

Tadelis, Steven (2002): Complexity, flexibility, and the make-or-buy decision. *American Economic Review*, 92(2), 433-437.

——— and Oliver Williamson (2013): Transaction Cost Economics. In: Gibbons and Roberts, pp. 159-189

Thompson, Edward P. (1967). Time, work-discipline, and industrial capitalism. *Past & Present* (38), 56-97.

Ure, Andrew (1835): *The Philosophy of Manufactures: or, An Exposition of the Scientific, Moral, and Commercial Economy of the Factory System of Great Britain.* London: Charles Knight.

Van den Steen, Eric (2010): Interpersonal Authority in a Theory of the Firm. *American Economic Review* 100(1), 466-90.

Weber, Lauren (2017): The end of employment. *Wall Street Journal*, February 17, 2017

Wernerfelt, Birger (1997): On the nature and scope of the firm: An adjustment-cost theory. *Journal of Business* 70 (4), pp. 489-514.

——— (2004): Governance of adjustments. *Journal of Business* 77 (S2), pp. S3-S24.

——— (2015). The comparative advantages of firms, markets and contracts: a unified theory. *Economica* 82(326), 350-367.

Whinston, Michael D. (2003): On the transaction cost determinants of vertical integration. *Journal of Law, Economics, and Organization* 19 (1), 1-23.

Williamson, Oliver E. (1971). The vertical integration of production: market failure considerations. *The American Economic Review*, 61(2), 112-123.

——— (1975): *Markets and Hierarchies.* New York (The Free Press).

——— (1985): *The Economic Institutions of Capitalism.* New York (The Free Press).

——— (1991): Comparative economic organization: The analysis of discrete structural alternatives. *Administrative Science Quarterly* pp. 269-296.

Zanarone, Giorgio (2012): Coase (1937) revisited: Endogenous fiat in firms and markets. *International Review of Economics*, 59(2), 201-221.

——— (2013): Contract adaptation under legal constraints. *The Journal of Law, Economics, & Organization*, 29(4), 799-834.