

## Interpersonal Authority in a Theory of the Firm

By ERIC VAN DEN STEEN\*

*This paper develops a theory of the firm in which a firm's centralized asset ownership and low-powered incentives give the manager, as an equilibrium outcome, interpersonal authority over employees (in a world with open disagreement). The paper thus provides micro-foundations for the idea that bringing a project inside the firm gives the manager control over that project, while explaining concentrated asset ownership, low-powered incentives, and centralized authority as typical characteristics of firms. The paper also leads to new perspectives on the firm as a legal entity and on the relationship between the Knightian and Coasian views of the firm. (JEL D23, L20)*

Interpersonal authority (i.e., the power of a superior to tell her subordinates what to do, with the reasonable expectation that they will obey) is a cornerstone of organization. It is most visibly expressed in the chain of command or in the managerial hierarchy. As Kenneth J. Arrow (1974) noted, “the giving and taking of orders...is an essential part of the mechanism by which organizations function.” In fact, employment is so closely linked with authority that people are often said to “become their own boss” when they become self-employed. Yet being an employee does not mean abandoning free will: the employee decides whether to obey the boss's directives. This raises the question of whether important institutional features of the firm might be motivated by the need to induce compliance.

The purpose of this paper is to develop a theory of the firm in which the firm is indeed a mechanism to give a manager interpersonal authority over employees. I will say that one person has interpersonal authority over another if: (i) the first person tells the second what to do, (ii) the second person tends to act in accord with these instructions, and (iii) he often does so against his own beliefs or immediate preferences (which distinguishes authority from advice). In particular, Herbert Simon (1947) stressed that “the characteristic which distinguishes authority from other kinds of influence is...that a subordinate holds in abeyance his own critical faculties...and uses...a command...as his basis for choice....Obedience...is an abdication of choice.” He also observed that “when the disagreement is not resolved by discussion, persuasion, or other means of conviction, then it must be decided by the authority of one or the other participant. It is this ‘right to the last word’...which is usually meant in speaking of ‘lines of authority’...”<sup>1</sup>

\*Harvard Business School, 15 Harvard Way, Boston MA 02163 (e-mail: [evandensteen@hbs.edu](mailto:evandensteen@hbs.edu)). John Roberts and Bob Gibbons introduced me to, and gave me most of my insights in, the theory of the firm. Both also contributed a lot to this paper, with Bob giving extensive feedback on multiple versions. I further thank Wouter Dessein, Mathias Dewatripont, Oliver Hart, David Scharfstein, Jordan Siegel, Ravi Singh, and especially Bengt Holmström for many conversations that helped me structure my thinking on these issues. I finally thank the participants in the SITE conference, the Tinbergen workshop on economics of the workplace, the Columbia economic theory workshop, the MIT organizational economics lunch, the Harvard-MIT organizational economics seminar, the Oliver E. Williamson Seminar on Institutional Analysis, the Stanford GSB-Economics seminar, the UCLA policy and strategy seminar, and the Princeton Behavioral Economics seminar for their comments and suggestions.

<sup>1</sup>This “interpersonal authority” differs from “decision authority” (i.e., the ability of a manager to make an interpersonal decision), which has been studied in some recent literature (Paul Milgrom and John Roberts 1988; Canice J. Prendergast 1995; Philippe Aghion and Jean Tirole 1997; Susan Athey and Roberts 2001; George Baker, Robert Gibbons, and Kevin J. Murphy 2006; Wouter Dessein, Luis Garicano, and Robert Gertner 2006), by its explicit focus on the interpersonal nature of authority and on the possibility of disobedience.

The idea that authority plays a central role in the nature and function of a firm has a long tradition among economists, including Frank Knight (1921), Ronald Coase (1937), Simon (1947, 1951), Arrow (1974), and Oliver Williamson (1975). Coase (1937), for example, likens the firm to a “master and servant” relationship. Armen A. Alchian and Harold Demsetz (1972), however, argued that “[the firm] has no authority...any different...from ordinary market contracting” and that “[the firm] can fire or sue, just as I can fire my grocer by stopping purchases from him or sue him for delivering faulty products.” In response, economists looking to build formal theories of the firm have focused on other foundations, ignoring authority. For example, the property rights theory (Sanford J. Grossman and Oliver D. Hart 1986; Hart and John Moore 1990; Hart 1995) defines a firm as a set of assets under common ownership but has no role for interpersonal authority.

In this paper, I show how a firm’s centralized asset ownership and fixed-wage contracts emerge endogenously as ways to give the firm’s manager, in equilibrium, interpersonal authority over the firm’s employees. I thus derive the following three observations as a bundle of equilibrium practices: (i) firms hire people under low-powered incentive contracts to work on the firm’s projects (Knight 1921; Simon 1951; Bengt Holmström and Milgrom 1994); (ii) a firm’s manager has interpersonal authority over these employees (Knight 1921; Coase 1937; Simon 1951; Arrow 1974; Williamson 1975); and (iii) firms own the assets that are necessary for their activities (Hart 1995; Holmström 1999).

To derive this result, I study a setting in which two players can start a pair of (vertically or horizontally) related projects. Each of the projects requires one asset and involves one decision, with decision externalities between the projects. Each player can execute only one project and only the player executing a project can make the corresponding decision. A key complication is that the players may openly disagree, in the sense of differing priors, on the right decisions. Decisions are not contractible, but the projects’ payoffs are. Moreover, each player can tell the other—through nonbinding cheap-talk—what he would like the other to do. Finally, cooperation is at will: people can walk away from the projects. This setup is careful not to postulate any changes (in contracting possibilities or informational conditions) from bringing a transaction inside the firm (Hart 1995; Gibbons 2005).

In this setting, two distinct but internally coherent bundles of practices emerge: an integrated or firm-like regime (at high decision externalities) and a nonintegrated or market-like regime (at low decision externalities). In the nonintegrated or market regime, neither player has authority over the other: each asset is owned by a different person who executes the corresponding project according to her own views and who gets all its residual income. In the integrated or firm regime, on the other hand, both assets are owned by one person who holds all residual claims on both projects and has—as an equilibrium outcome—interpersonal authority over the other person.

The starting point for the intuition is the fact that it is optimal—with differing priors—to collocate residual income and control (Van den Steen, forthcoming): as a person gets more control rights, she values income rights more highly (since she believes that she makes better decisions than others), so it is optimal to give her more income rights; as a person gets more income rights, she values control rights more highly (since she cares more and believes that she makes better decisions), so it is optimal to give her more control rights. To move control rights around, the players write (as an equilibrium outcome) efficiency-wage type contracts that create interpersonal authority for one player over the other. The role of low-powered incentives is then, first, to minimize the employees’ temptation to disobey when they disagree with their boss (Van den Steen 2007a) and, second, to collocate income and control. The role of centralized asset ownership is, first, to affect the *level* of the outside options in a way that makes the employee obey the manager and, second, to minimize the constraints that asset ownership imposes on the allocation of income (through the requirement that a player’s continuation value exceed his assets’ outside

value). The idea behind the obedience effect is that moving asset ownership from the agent to the principal lowers the outside option of the agent and raises that of the principal, making it more costly for the agent to get fired and easier for the principal to commit to firing a disobeying agent. The theory applies to any asset that incurs an opportunity cost from being used in the project. The results do not disappear if assets can be acquired or traded on the open market. While the results for low-powered incentives depend on open disagreement, the results for asset ownership seem to apply more broadly. I also show that the same results obtain in a model with a mix of both disagreement and either (limited) private benefits from decisions or moral hazard in the form of unobservable effort. The efficiency wage then overcomes the temptations to disobey from both differing priors and private benefits or costs. I finally argue that this model is closely related to Knight's (1921) theory of entrepreneurship.

In terms of firm boundaries, the force for integration is the decision externalities. The force against integration comes in two forms. In the main model, nonintegration is driven by the players' differential confidence about different decisions, which can make it optimal for each player to "own" the project for which he is the most confident player. Nonintegration then allocates uncertainty to the person who is most confident about its resolution. In the model with moral hazard, on the other hand, it is the need for pay-for-performance incentives that drives nonintegration, since the incentives both tempt the employee to disobey orders when he disagrees and make it optimal to shift control to the employee. The firm boundaries then result from a trade-off between cooperation and motivation (Roberts 2004). I further show that firm boundaries may matter even when actions are perfectly contractible, which suggests a new perspective on the relationship between the Coasian and Knightian views of the firm. Somewhat contrary to the view of integration as a substitute for contracts, I also show that improved contractibility may actually lead to more, rather than less, integration.

The theory in this paper is formulated—following Knight (1921) and Coase (1937)—in terms of a manager-owner or entrepreneur-owner. This obviously raises the question whether the model can be used to study large firms with shareholders. The key to answering that question is the underlying conception adopted here of a firm as a legal person, which thus has all the standing of a real person (e.g., it can own assets and write contracts) but which is distinct from its shareholders (e.g., no shareholder directly owns any asset) and whose powers are exercised by a manager.<sup>2</sup> The key insight then is that the firm as a legal person aggregates asset ownership and thus allows the firm's manager, when exercising the firm's powers, to act as the "as if" owner of all assets and as the "as if" party to all contracts and thus claim the authority that goes with such centralized ownership. But, at the same time, the firm as a legal entity preserves the benefits of dispersed ownership for risk diversification and weakening of budget constraints. It thus allows the best of both concentrated and dispersed asset ownership and contracts. I will discuss later how it also changes the perspective on firm boundaries. However, since the current paper considers only entrepreneur-owned firms that do not change hands, the firm (as a legal person) and the owner (as a physical person) are interchangeable, so that I will not distinguish them formally here. Van den Steen (2007b) develops these ideas on the role of the firm as a legal entity more formally.

### A. Contribution

This paper makes three contributions. First, it develops a theory of the firm in which the firm is a mechanism to give the manager interpersonal authority over the firm's employees. Second, it shows, in the process, how asset ownership may convey interpersonal authority and

<sup>2</sup> While being a legal person is a legal property of a firm, it has distinct implications in this context.

thus determine effective control. Finally, the paper shows how differential confidence and the motivation-cooperation trade-off may drive firm boundaries, and that firm boundaries may matter even when actions are contractible. All three contributions, however, are about one thing: how the firm is a mechanism to give the manager interpersonal authority over employees.

### B. Literature

To relate this paper to the literature on the theory of the firm, it is useful to categorize that literature into two groups. The first comprises contributions in which interpersonal authority plays no role, such as Alchian and Demsetz (1972), the Grossman-Hart-Moore property rights theory, Raghuram G. Rajan and Luigi Zingales (1998), Hart and Holmström (2002), Hart and Moore (2008), and Jonathan Levin and Steve Tadelis (forthcoming). Relative to this group, I show how the firm, as an institution, may generate interpersonal authority. Moreover, while these theories define “the firm” as a label for a set of contracts (including property rights), I conceptualize the firm instead as a legal entity.

The second group comprises contributions that take interpersonal authority to be a key aspect of firms, such as Knight (1921), Coase (1937), Simon (1951), or Williamson (1975). While these contributions clarify important aspects of authority, they do not address the question why (or when) employees would be more likely than contractors to obey, and how that relates to firm characteristics. For example, Williamson’s (1991) argument on forbearance and Holmström and Milgrom (1994) are essentially about the *need* to rely on authority instead of *why* an employee would actually obey. The current paper differs from this second group by the fact that I formally derive the firm as *a mechanism to generate* interpersonal authority. In that sense, the paper is complementary to this literature, since it provides potential formal microfoundations for some of its arguments and assumptions. The Web Appendix (available at <http://www.aeaweb.org/articles.php?doi=10.1257/aer.100.1.466>) actually details how this paper is complementary to most of the perspectives on the firm outlined in Gibbons (2005).<sup>3</sup> But I also derive new insights on firm boundaries: I show how both differential confidence about the central decisions for a project and the trade-off between effort and cooperation may drive firm boundaries and how firm boundaries may matter even when actions are perfectly contractible.

Of particular interest here is Holmström and Milgrom (1994), who argued that a firm is characterized by a similar set of interrelated practices: low-powered incentives, employees not owning assets, and the *use* of authority to exclude employees from certain returns from outside projects. Apart from building on very different mechanisms, the current paper differs in two respects. First, it is about the source (instead of the use) of interpersonal authority. Second, it explains not only why employees don’t own assets, but also why it should be the firm (rather than a third party) that owns the assets. This is discussed in more detail in the Web Appendix. I build on this work by looking at the firm as a set of interrelated practices and by considering low-powered incentives as a purpose, rather than an unfortunate consequence, of transacting through a firm.

This paper is also related to work that studied organizations through the lens of efficiency wages, such as Guillermo A. Calvo and Stanislaw Wellisz (1979) and Yingyi Qian (1994), who study how efficiency wages for effort affect hierarchies, Van den Steen (2007a), which is discussed below, and Anthony Marino, John Matsusaka, and Jan Zábajník (forthcoming), who study how organization and market characteristics affect obedience and thus the equilibrium allocation of control. None of these papers, however, considers the role of assets, the theory of the firm, or firm boundaries.

<sup>3</sup> Firms are very complex institutions that cannot be understood from just one perspective to the exclusion of others. For interesting perspectives on this richness, see also Holmström and Roberts (1998).

There is also a related literature on relational contracts. Birger Wernerfelt (1997) informally argues that a manager has authority over employees through a relational contract, but without an argument of how such contracts differ from relational contracts in markets. Baker, Gibbons, and Murphy (2002) and Levin (2003) have other similarities: the basic issue is unenforceability of actions and payments; the main threat is termination of a relationship; low-powered incentives may be optimal; and assets play a role (in BGM). But there are equally important differences: both the low-powered incentives and the asset ownership are driven by the risk of renegeing on compensation for effort (which plays no role here); the relationship between incentives and assets runs causally in the opposite direction; there is no issue regarding how the principal commits to firing (which matters here); asset ownership (in BGM) works through ex post renegotiation (as in the property rights theory but unlike here); and, most important of all, there is no interpersonal authority or abdication of choice in these relational contracting models. Note that, while the current paper's results on asset ownership (and low-powered incentives) could probably be derived in a relational contracting setting, they do not require a repeated-game structure.

The idea that asset ownership is related to authority has been suggested before, but in a different sense than the interpretation or formalization in this paper. Hart (1995) shows that asset ownership makes others orient their specific investments toward the asset owner, which he interprets as asset ownership conveying authority, but which is very different from the issues of orders and obedience. Holmström (1999) argues that "asset ownership conveys the CEO...the ability to restructure the incentives," which is a form of decision authority (over incentives), but not interpersonal authority. Benjamin E. Hermalin (1999) argues informally that centralization of control rights may prove difficult without also centralizing asset ownership: if employees own assets, they can force decisions by threatening hold-up or exit. His argument is thus essentially the reverse of part of the argument in this paper. Wernerfelt (2002), finally, shows that the person in control should own the assets since he can better internalize the effect of his decisions on the assets.

The current paper builds on Van den Steen (2007a, forthcoming). The first showed, among other things, that interpersonal authority will go together with low-powered incentives, while the second showed, among other things, that authority and residual income should be colocated when people have differing priors. Both are important elements in the current paper. Relative to these papers, the current paper adds asset ownership as a lever and, more important, shows how these elements combine into a coherent theory of the firm that is intimately linked with interpersonal authority and that generates new insights for firm boundaries. Apart from building on these papers, I apply Van den Steen (2008), which shows how Coase (1960) is affected by open disagreement, to show how firm boundaries may matter even when actions are contractible.

## I. The Model

The model in this paper captures a setting in which two persons can start a pair of related projects, such as microprocessors and PCs, or game consoles and games. Each project requires one asset and involves one decision, with decision externalities between the two projects. A key complication is that the two persons may openly disagree on the right decision. The question is, then, how the assets will be owned, who will (effectively) make each decision, and how the returns will be shared.

Formally, consider two projects ( $R_1$  and  $R_2$ ), two assets ( $a_1$  and  $a_2$ ), and two players ( $P_1$  and  $P_2$ ). A project is formally defined as a revenue stream ( $R_k$ ) that requires for its realization the asset  $a_k$  and the full-time involvement of one player. Due to the need for full-time involvement, each player can execute only one of the two projects. Let  $P(k) \in \{P_1, P_2\}$  denote the player who executes project  $R_k$ . To execute  $R_k$ , that player  $P(k)$  will have to make a decision  $D_k \in \{X_k, Y_k\}$ . One and only one of these choices is correct, as captured by the state variable  $S_k \in \{X_k, Y_k\}$ .

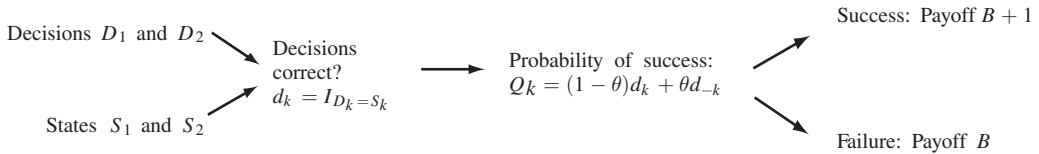


FIGURE 1. ACTIONS AND PAYOFFS

As depicted in Figure 1, each project will be either a failure or a success, with respective payoffs  $B > 0$  and  $B + 1$ . The probability of success, denoted by  $Q_k$ , depends on which decisions are correct. In particular, let  $d_k = I_{\{D_k=S_k\}}$  be the indicator that decision  $D_k$  is correct and analogous for  $d_{-k}$  (where  $-k$  indexes the project other than  $R_k$ ) and let  $Q_k = (1 - \theta)d_k + \theta d_{-k}$  with  $\theta \in (0, 0.5)$ . In other words, a project's success depends mainly on the project's decision (and more so as  $\theta$  decreases), but there are decision externalities when  $\theta > 0$ .<sup>4</sup>

The states  $S_k$  are unknown, but each player  $P_i$  has a subjective belief  $\mu_k^i$  that  $S_k = X_k$ . A key assumption is that (it is common knowledge that) players have differing priors, i.e., they can disagree about  $S_k$  even though neither has private information. I will discuss this assumption in more detail at the end of this section. Since the players may have differing priors but have no private information about  $S_k$ , they will not update their beliefs when they meet someone with a different belief: they simply accept that people sometimes disagree.

To keep the analysis simple, both players' beliefs will be independent draws from a commonly known binary distribution: for some fixed  $\nu_k^i \in (0.5, 1)$ ,  $\mu_k^i$  equals  $\nu_k^i$  or  $(1 - \nu_k^i)$  with equal probability. This means that  $P_i$  believes half the time that  $X_k$  is the best course of action, and half the time that  $Y_k$  is the best course of action, but  $P_i$  always has the same confidence (or strength of belief in what he believes is best)  $\nu_k^i = \max(\mu_k^i, 1 - \mu_k^i)$ . Moreover, the two players disagree half the time. For simplicity, I will also assume that the  $\nu_k^i$ , which are given exogenously as parameters of the model, can take only two values,  $0.5 < \underline{\nu} \leq \bar{\nu} < 1$ .

An important element in the model is that players have alternative uses for the assets: if  $R_k$  is not executed, then the asset  $a_k$  can potentially be deployed elsewhere and that outside option accrues to the asset owner. Let  $\underline{u}_k^i$  denote  $P_i$ 's outside option when  $R_k$  is not executed (which may change over time). One important interpretation of these outside options—as studied in the working paper Van den Steen (2007c)—is that they represent an opportunity to execute the same project with a different player at a later date, such as hiring a new employee after firing an employee. Since the outside option derives from asset ownership,  $\underline{u}_k^i = 0$  if  $P_i$  does not own  $a_k$ . If  $P_i$  owns  $a_k$ , then  $\underline{u}_k^i$  starts at some  $\underline{u} > 0$  but decreases over time as alternatives disappear. This decrease over time will be detailed in the timeline. I will assume that assets are sufficiently important, in particular that  $\underline{u} > B$ . Assume also that prior to the game, the assets are randomly allocated to players. This starting asset allocation does not affect the outcome of the game (except for an up-front transfer).

I now discuss the timing of the game (Figure 2). First, the players negotiate the asset ownership, the project allocation, and contracts for each project. This negotiation follows symmetric Nash bargaining with an immediate financial transfer and with the  $\underline{u}_k^i$  as outside options. The project allocation simply matches projects to players (one to one), which determines who will execute what project and thus make what decision. The project contracts determine the allocation of income for each project. In terms of contractibility, I will assume that the decisions  $D_k$

<sup>4</sup>I exclude  $\theta = 0$  to simplify the analysis, but the results would hold for that case.

1	2	3
Contracting a. The players negotiate asset ownership, project allocation, and project contracts. b. The beliefs $\mu_k^i$ get (privately) drawn.	Orders and decisions a. Each player $P_i$ can send a cheap-talk message from $\{X_k, Y_k\}$ . b. Each player $P(k)$ publicly chooses his action from $\{X_k, Y_k\}$ . With probability $p$ , the outside options drop to $\underline{u}_k^i = 0$ . c. Each participant in a project can terminate that project.	Execution and payoffs a. Assets are committed (so that $\underline{u}_k^i = 0$ for sure). Success or failure is revealed. b. Each participant in a project can terminate that project. c. The contract terms $(w_k^i, \alpha_k^i)$ get executed.

FIGURE 2. TIME LINE OF BASIC MODEL

(in stage 2b) are not contractible, and—once people are assigned to projects—neither is the right to make these decisions: the person who executes project  $R_k$  automatically makes decision  $D_k$ . I will assume, however, that each project's payoff is contractible. A contract for project  $R_k$  will then consist of a fixed payment  $w_k^i$  and a share  $\alpha_k^i \in [0, 1]$  of the extra project revenue upon success. For budget balance reasons, for each project we need  $w_k^1 + w_k^2 = B$  and  $\alpha_k^1 + \alpha_k^2 = 1$ . The assumption  $\alpha_k^i \in [0, 1]$  says that players will not take on more joint uncertainty than what is necessary to allocate the project's revenue. It prevents players from using the project's outcome as the basis for a pure bet.<sup>5</sup> One way to derive this condition endogenously is to give each player the ability to sabotage the project, i.e., the ability to make sure that the project fails (Van den Steen 2007a), for example, by withholding costless but necessary effort. In that case, any contract with  $\alpha_k^i \notin [0, 1]$  would give one of the players a strict incentive to sabotage that project. Anticipating that, the other would never accept the "bet." In fact, if players could bet on the state in some other setting (such as Wall Street or Las Vegas), then even the tiniest probability of sabotage would endogenously lead to  $\alpha_k^i \in [0, 1]$ . Other ways to endogenize this constraint include risk-aversion (Michael Harrison and David M. Kreps 1978), ambiguity aversion (in this case, discounting payoffs about which there is open disagreement), capital constraints, or adverse selection with respect to private information (Stephen Morris 1994). For generality and simplicity, I impose the condition here as an assumption. In stage 1b, the players' beliefs get privately realized.<sup>6</sup>

In stage 2a, each player can, for each project  $R_k$ , send a cheap talk message from the set  $\{X_k, Y_k\}$  to the other player. (As will become clear in Section II, in equilibrium, at most one party—the boss—will send a cheap-talk message and it will be interpreted by the other party—the subordinate—as a nonbinding order. In the setup, though, both players are symmetric.) In order to focus only on interesting cheap talk equilibria, I will assume that (i) when indifferent, players prefer not to send any message (as if there is an infinitesimal cost  $c \downarrow 0$  of sending a message); (ii) when indifferent (but sending a message), they prefer to communicate their true belief  $Z_k^i$  (as if there is an infinitesimal cost  $l \downarrow 0$  from "lying"); and (iii) players select the Pareto-dominant message-equilibrium if one exists.

In step 2b, the players publicly choose their actions. Since the players' decisions are noncontractible, each agent chooses the decision that is best from his perspective, given his beliefs and the contract. (In this sense, an order from the boss does not directly constrain the subordinate's behavior.) The action choice sometimes goes together with a decrease in the outside

<sup>5</sup> Betting itself is not problematic, as long as it happens outside the context of the project.

<sup>6</sup> Drawing the beliefs after the contract negotiation captures the idea that important contentious issues may become clear only after the project is started. Imagine that there is an infinite number of potential issues (on which each player has fixed beliefs), and stage 1b reveals which of these issues is the relevant one.

options. This captures, for example, that the actions may be irreversible or, if the outside option consists of matching with another player, that it may become impossible to find another partner in time to execute the project. Formally, with probability  $(1 - p) \in (0, 1)$ , the outside options drop to  $\underline{u}_k^i = 0$ , while with complementary probability they remain unchanged at  $\underline{u}$ .<sup>7</sup>

After the decisions, each participant in a project has two opportunities to terminate that project, where a “participant in project  $R_k$ ” is defined to be any player  $P_i$  such that either  $P_i$  owns  $a_k$ , or  $P_i = P(k)$ , or  $(\alpha_k^i, w_k^i) \neq (0, 0)$ . (As will become clear in Section II, the opportunity to terminate a project will allow one player to effectively “fire” the other, although it will only take on that meaning in equilibrium. At this point, both players are symmetric.) The two opportunities correspond respectively to an early stage of the project—with the project’s outcome unknown but still some outside options remaining—and a late stage of the project—with the outcome known but no outside options. In particular, in step 2c—which represents the early stage—each participant has a first opportunity to terminate each project. In step 3a, the project enters the late stage: the assets are completely committed or used up, which means that the outside options for all players drop to zero for sure ( $\underline{u}_k^i = 0$ ), and the states  $S_k$  are revealed. In step 3b, each participant gets another chance to end the project. If a participant ends a project, then each  $P_i$  gets  $\underline{u}_k^i$  for that project and the project is over. In period 3c, finally, the payoffs are realized and the remaining contract terms  $(w_k^i, \alpha_k^i)$  are executed (for any project that has not been terminated).

For the outside value  $\underline{u}$ , I will also assume that  $B + \underline{v} > \underline{u} > (1 - \theta)(1 - \underline{v}) + \theta\underline{v}$ . The first inequality ensures that the project always gets executed in equilibrium, while the second ensures that a player can always commit to quitting when the other takes the decision that the first player thinks is wrong (disobeys). Both simplify the analysis considerably although the essential results would hold without the assumption. I will also assume that when payoff-indifferent on the actions, each player chooses the action that she thinks is most likely to succeed (as if each player gets a private benefit  $\gamma \downarrow 0$  from success) and that each participant, when indifferent, stays in stages 2c and 3b. Finally, I will focus on pure-strategy subgame-perfect equilibria that are not Pareto dominated.

To summarize, there are essentially two sets of moving parts in this model, each with a clear purpose. The first set enables interpersonal authority to play a role: the drawing of beliefs (so that there is something to give orders about); the cheap talk messages (so that an order can be given); and the decisions (so that the other can either obey or disobey). The second set enables the efficiency wage: the contracting up-front (so that they can agree on a wage); the ability to quit (so that the principal can fire the agent); and the outside options.

### A. The Differing Priors Assumption

The model assumes that people can openly disagree, i.e., they have differing priors.<sup>8</sup> While Section IIC considers how this assumption affects the results, I discuss it here from a more general perspective.

Note first that the differing priors assumption has a long tradition with, among others, Arrow (1964); Robert Wilson (1968); Harrison and Kreps (1978); Hal R. Varian (1989); Morris (1994, 1997); Jose A. Scheinkman and Wei Xiong (2003); Muhamet Yildiz (2003); Van den Steen (2004); Markus K. Brunnermeier and Jonathan A. Parker (2005); Arnoud W. A. Boot, Radhakrishnan

<sup>7</sup> The uncertainty creates a temptation to disobey, since the other player may have no recourse.

<sup>8</sup> The assumption that players have no private information is made for analytical convenience. If players also had private information, they would update their beliefs, but disagreement would remain (Morris 1997). For a more conceptual discussion of differing priors, see Morris (1995). Note that differing priors is *not* the same as private information that is impossible to communicate.

Gopalan, and Anjan V. Thakor (2006); and Luigi Guiso, Paola Sapienza, and Luigi Zingales (2006). There has been a rapid rise in recent years, in part due to the growing popularity of behavioral economics, which often implicitly assumes differing priors. There is also a burgeoning empirical literature, such as Joseph S. Chen, Harrison Hong, and Jeremy C. Stein (2002) and Augustin Landier and David Thesmar (2009). Furthermore, Hong and Stein (2007) argue that “disagreement models... represent the best horse on which to bet [as the future consensus model for behavioral finance].”

The assumption of (originally unbiased) differing priors captures the fact that people may have different “mental models” or “belief systems,” or different intuition which leads people with identical data to draw different conclusions. Consider, for example, a manager’s belief whether a particular person or group of people is trustworthy and how that may influence her decision whether to do business with that person or group. This kind of issue is common in organizations. Indeed, the fundamental role of “belief systems” or “mental models” in organizations has been stressed by academic studies of managers and managerial decision making (Gordon Donaldson and Jay W. Lorsch 1983; Edgar H. Schein 1985).

A natural question is why players don’t simply discuss until they reach agreement or why firms don’t simply hire employees with the “right” beliefs. In both cases, the choice is essentially a time and cost trade-off, and in many cases authority is the right option. Imagine, for example, the deadlock if a CEO (or a dean) needed to persuade all her subordinates before making any decision. Analogously, sorting on beliefs is often expensive, since beliefs are difficult to extract while turnover destroys organizational and firm-specific human capital. This implies that firms are more likely exactly when authority is useful.

A final question is where such differing priors would come from in a Bayesian framework. There are two ways to think about this. Since the prior for this game is a posterior from earlier updating, bounded rationality (of which the player is not aware) will often lead to differing priors, even when starting from a common prior. Unconsciously forgetting some of the data used to update beliefs, for example, would do. A second—more philosophical and more controversial—argument is that people may be born with differing priors: in the absence of information there is no reason to agree, and priors are just primitives of a model. I am agnostic about the source of the disagreement; I just explore its consequences.

## II. Interpersonal Authority, Ownership, and Contracts

This section derives and discusses the results of the paper. Section IIA derives the main proposition and gives the basic underlying intuition. To clarify some important issues that arise from this analysis, I show that the results also hold in a model with (limited) private benefits, and further study, in Section IIB, a variation with moral hazard. Apart from clarifying the mechanisms and intuition, the moral hazard model also gives insight in the trade-off between motivation and cooperation (Roberts 2004). Section IIC discusses some further issues and implications.

### A. The Main Analysis

The main result of the paper is captured in Proposition 1 below, which shows that there are two equilibrium regimes: an integrated or firm-like regime at high decision externalities (high  $\theta$ ), and a nonintegrated or market-like regime at low decision externalities. In the nonintegrated or market regime, no player has authority over the other: each asset is owned by a different person who executes the project according to her own views and who is its residual claimant. In the integrated or firm regime, on the other hand, both assets are owned by one person who holds all residual claims and has interpersonal authority over the other person. Moreover, as a side result,

I also show that it is the most confident person (i.e., the one with the higher  $\nu_k^i$ ) who becomes the owner-manager in the integrated regime.

For the formal statement of the proposition, I will use the following definitions. Let  $Z_k^i$  denote the action that  $P_i$  believes is most likely to be correct for decision  $D_k$ , and let  $-i$  index the player other than  $P_i$  and  $-k$  index the project other than  $R_k$ . I will say that the equilibrium of the subgame starting in 2a displays “no authority” if neither player sends a cheap talk message in 2a; each player  $P_i$  (executing  $R_k$ ) just chooses  $Z_k^i$  in 2b; and neither player quits. I will say that the equilibrium is “authority by  $P_i$ ” if  $P_i$  (executing  $R_k$ ) sends a cheap talk message  $Z_{-k}^i$  telling  $P_{-i}$  what to do (while  $P_{-i}$  sends no cheap talk message);  $P_i$  chooses  $Z_k^i$  while  $P_{-i}$  does what  $P_i$  told him to do;  $P_i$  quits (at least one project) in 2c if and only if  $P_{-i}$  did not choose  $Z_{-k}^i$  (but never quits any project if  $P_{-i}$  chooses  $Z_{-k}^i$ ); and neither player quits in step 3b. Let  $\hat{O}$  be the set of parameters that satisfy the assumptions of Section I, which is nonempty, and let  $\hat{\Omega}$ , with typical element  $\omega$ , be the parameters excluding  $\theta$  such that there exists  $(\omega, \theta) \in \hat{O}$ .

**PROPOSITION 1:** *The parameter space  $\hat{O}$  is partitioned into two sets,  $F_I$  and  $F_{NI}$ , which have respectively an integrated (firm) equilibrium and a nonintegrated (market) equilibrium.*

- (i) *In  $F_I$ —the integrated (firm) set—both assets are owned by one player, say  $P_i$ ; that asset owner  $P_i$  is the full residual claimant on both projects ( $\alpha_1^i = \alpha_2^i = 1$ ); the other player  $P_{-i}$  gets no residual income but a (strictly positive) wage  $w_k^{-i} > 0$  on one or both projects; and the subgame equilibrium starting in stage 2a is “authority by  $P_i$  over  $P_{-i}$ .”*
- (ii) *In  $F_{NI}$ —the nonintegrated (market) set—each asset is owned by a different player, say  $a_1$  by  $P_i$  and  $a_2$  by  $P_{-i}$ ; each player executes the project for which he owns the asset; each player gets all residual income from “her” project (i.e.,  $\alpha_1^i = \alpha_2^{-i} = 1$ ); and the subgame equilibrium starting in stage 2a is “no authority,” i.e., no player has authority over the other.*

*The equilibrium goes from nonintegrated to integrated as the decision externalities increase: for any  $\omega \in \hat{\Omega}$ , there exists a  $\hat{\theta}$  such that  $(\omega, \theta) \in F_{NI}$  if  $\theta \leq \hat{\theta}$ , while  $(\omega, \theta) \in F_I$  when  $\theta > \hat{\theta}$ .*

*If there exists a player  $P_i$  such that  $\nu_k^i \geq \nu_k^{-i}$  for all  $k$  with strict inequality for at least one  $k$ , then that player  $P_i$  will have authority in the firm partition  $F_I$ , i.e., the player with most confidence will be in charge. In  $F_{NI}$ , each asset will be owned by the player who is most confident about the respective project’s decision.*

**PROOF:**

See Appendix A.

This result is driven by a combination of forces—some that were derived earlier and some that are new to this paper—which arise when players disagree on a project that requires assets and that depends on incontractible decisions.

A known effect is that differing priors make it strictly optimal to concentrate all income and control rights of a project in one hand (Van den Steen, forthcoming). This brings the focus on the mechanisms that allow players to move income and control around or that constrain such movements.

Consider first control rights. The basic mechanism in this paper to move control rights around is the use of an efficiency wage: in equilibrium, one player gets a high wage while the other tells him what to do and (credibly) threatens to fire him if he disobeys.

There are two mechanisms in this model that complement the efficiency wage. The first mechanism—derived in Van den Steen (2007a)—is that by giving the agent low-powered incentives, such as a fixed wage, the principal minimizes the agent’s temptation to disobey the principal when the two of them disagree on the right course of action: since low-powered incentives make the agent care little about success or failure of the project, the agent will rather make a decision that is wrong (from his subjective perspective) than risk getting fired and losing the efficiency wage.<sup>9</sup> Outcome-contingent pay thus requires a higher efficiency wage (as counterweight) to maintain compliance. When paying an efficiency wage is costly or restricted, lower-powered incentives may be optimal.

The second mechanism to strengthen the manager’s interpersonal authority, which is new to this paper, is through the allocation of asset ownership. Asset ownership determines who gets the assets when cooperation breaks down and thus determines the *level* of each player’s outside option. These levels determine the cost of firing and getting fired, and thus the agent’s incentives to obey the principal. Moving an asset from an agent to the principal reduces the agent’s outside option and thus increases his cost of getting fired, while simultaneously increasing the principal’s outside option, making it easier for her to commit to firing. Both effects strengthen the principal’s authority.

To make the intuition for these mechanisms more formal, consider the condition (from the proof of Proposition 1) for  $P_2$  to obey when  $P_2$  executes  $R_k$ , and  $P_1$  has committed to quit  $R_k$  if and only if  $P_2$  disobeys:

$$w_k^2 > \underline{u}_k^2 + \alpha_k^2 \left( \frac{(1-p)}{p} (1-\theta)(2\nu_k^2 - 1) - (1-\nu_k^2) \right) + \alpha_{-k}^2 \frac{1}{p} \theta(2\nu_k^2 - 1).$$

The wage  $w_k^2$  is the efficiency wage, and the right-hand-side expression is the minimal efficiency wage level to make  $P_2$  obey. For the result on assets, note that the right-hand side is increasing in  $P_2$ ’s outside option  $\underline{u}_k^2$ . In fact, when all residual income is allocated to  $P_1$ ,  $P_2$  will never obey if he owns the asset. For the result on low-powered incentives, note that if  $p$  is sufficiently small, then the minimal efficiency wage is increasing in both  $\alpha_1^2$  and  $\alpha_2^2$ : lower-powered incentives facilitate authority.<sup>10</sup> A third result, however, is that it may sometimes be impossible to find a set of contractual values that satisfy this condition. In that case, nonintegration is the only alternative. While that happens only off the equilibrium path in the basic model, it will be a direct determinant of integration in the variation with moral hazard in Section IIB. Finally, to see the other side of asset ownership, note that  $P_1$  will be committed to quit upon disobedience if and only if  $\underline{u}_{1,k} > w_k^1 + \alpha_k^1 ((1-\theta)E^1(d_k) + \theta E^1(d_{-k}))$ , so that only a player who owns an asset can commit to firing the other player.

Before considering income, there are some issues that deserve explicit attention. A first issue is the role of the (second) opportunity to terminate the project, in stage 3b. The ability to end the project at that time effectively restricts the wage (to  $w_k^i \leq B$ ): any wage promise above  $B$  will be voided upon failure since at least one player will terminate the project. Without such implicit wage limit, obedience could always be enforced by promising an arbitrarily high wage and threatening to fire upon disobedience. The restriction can be motivated both by the fact that such extremely high wage promises are often legally invalid (when they exceed the value of the project) and by the fact that they create a motivation for one party to get out of the contract. The

<sup>9</sup> Note that I use “incentive” (in “low-powered incentives”) in the sense of an outcome-based contractual payment, as in “a wage based on the number of units produced by a factory pieceworker” (Merriam-Webster unabridged).

<sup>10</sup> To understand the need for low  $p$ , note that the average residual income acts in itself as an efficiency wage that the agent loses when the project is terminated. A large  $p$  increases the role of the average level relative to its outcome dependence. This force does not work in the opposite direction (i.e., toward higher  $\alpha_k^2$  when  $p$  is relatively high): instead, this compliance condition will simply never be binding when  $p$  is large.

essential results of the paper would hold whenever it is costly to guarantee a high wage. For example, the results would hold if a player could exert (privately costly) effort in 3b to get out of the contract, even if the player's probability of succeeding were tiny: preventing the wasteful private effort would by itself be a reason to avoid high wages, and thus to allocate ownership according to the proposition. A second issue that deserves clarification is the role of the assumption that assets are sufficiently important ( $\underline{u} > B$ ). Without this assumption, asset ownership would at times not matter, since one player could "rent" the asset from the other for a fixed payment, and that promise would be credible since the project has sufficient guaranteed income  $B$  to compensate for the value of the asset. Third, the role of the assumption that control rights cannot be traded is to ensure that interpersonal authority can really play a role. Finally, there is also a nice way to capture the notion of obedience as "abdication of choice" in this model: if observing your own belief would require some infinitesimal effort or cost, then employees in  $F_1$  would (in equilibrium) not bother to learn their own beliefs and blindly follow orders. They would thus literally "abdicate choice."

Apart from moving around control, players will also want to move around income. This might seem trivial since the model makes income contractible. Asset ownership, however, does impose implicit constraints on the allocation of income. In particular, since a player can always quit and collect his outside option, a player's expected continuation income from a project must exceed the current outside options from his assets that are used in the project. When a project's fixed income  $B$  is insufficient to cover these outside options, then asset owners will have to get a share of the residual income. Unlike a common priors context, a reallocation of residual income may directly affect efficiency in this differing priors setting. This provides another reason, beyond the need to induce obedience, why asset ownership and authority coincide: the asset owner will often necessarily become a (partial) residual claimant, and since it is optimal to collocate residual income and control, it will then also become optimal to collocate residual control and asset ownership. This link between asset ownership and authority is, to my knowledge, new.

Note that this effect of outside options on the allocation of income is very different from that in Grossman and Hart (1986) and similar work, where, the outside options affect only the transfer only at the time of bargaining, which is actually completely inconsequential for the results of the current paper. It is, instead, the role of assets as a continuing outside option after the contracting that constrains any promises for future transfers and that enforce compliance.

The firm boundaries in the model are in part driven by the result of Van den Steen (forthcoming) that income and control should be concentrated as much as possible with the most confident person. In particular, the combination of this result with the earlier one immediately generates a tension between integration and nonintegration: on the one hand, the decision externalities make it optimal to concentrate all income and control for *both* projects in the hands of *one* person; on the other hand, however, different players being more confident about different decisions can make it optimal to give each decision, and the most affected income, to the player who is most confident about that decision. This logic is reflected in the proposition: nonintegration happens in equilibrium when different players are most confident about different decisions and when the externalities are limited. I will discuss later how a similar tension may arise from the combined need for private effort and for interpersonal authority.

With respect to the Alchian and Demsetz (1972) critique, this theory thus differentiates a grocer from an employee based on both asset ownership and income.<sup>11</sup> Relative to an employee, a grocer will be less inclined to obey my order (e.g., to paint his grocery purple) since (i) he bears

<sup>11</sup> A customer may well have some interpersonal authority over a grocer. This paper's argument is that a boss typically has *more* authority over an employee (especially for particular decisions).

all the business consequences of following my order, and (ii) he retains his grocery and *at most* loses me as a client. Relative to that of a boss, my threat is typically less credible since (i) I often have no alternative to this grocery if I don't own it, and (ii) I have very little at stake in the color of the grocery and thus little incentive to follow up on my threat if doing so is costly to me. Note that employees are especially more likely to obey when the conflict is one of disagreement rather than private benefits or costs. This suggests that firms are most useful when it is more important to make the right decisions than to counter private effort or private benefits.

While these mechanisms are quite intuitive, the specific equilibrium outcome that  $\alpha_k^{-i} = 0$  in  $F_I$  raises two potential issues. The first is that the  $\alpha_k^{-i} = 0$  may give the impression that the result hinges on  $P_{-i}$  being completely indifferent about his own decisions and therefore willing to do whatever, including obeying  $P_i$ . This would be a very fragile result, since any exogenous interest of the worker, such as intrinsic motivation or private benefits, would break the result. To make very clear that this is *not* the case, I will do two things. First, I will show immediately that adding limited private benefits from the actions does not affect the outcome. Second, the next subsection introduces traditional moral hazard, i.e., private effort, into the model and shows that the key results are preserved. But complete indifference was in fact already broken in this basic model by the assumption that, when indifferent, the agent strictly prefers to follow his beliefs.

The second issue related to  $\alpha_k^{-i} = 0$  in  $F_I$  is the question of why asset ownership would matter if the players are close to indifferent. The answer to this question is twofold. First, the asset allocation does matter for authority, even when  $\alpha_k^{-i} = 0$ , for two reasons, both of which are linked to the fact that—due to the indifference assumptions—the employee will just follow his beliefs if he does not fear being fired. On the one hand, the principal would be unable to commit to firing the employee upon disobedience unless she has a good outside option, i.e., unless she owns the asset. On the other hand, if the employee owns the asset, then firing would not be a punishment since the outside option of the employee would be better than his wage (given that  $\underline{u} > B \geq w_k^{-i}$ ). A second reason why asset ownership matters even when  $\alpha_k^{-i} = 0$  is the role of assets as a constraint on the allocation of residual income. In particular, if  $\underline{u} > B$  then at least some residual income must be allocated to the owner of the asset, and that may be suboptimal in this differing priors context.

*Private Benefits from Actions.*—I will now show that the results of Proposition 1 are robust to adding a moderate amount of private benefits from undertaking specific actions. Assume that, for each of the projects, the player executing project  $R_k$  gets—apart from her share of the project revenue—a private benefit  $b \geq 0$  from one of the two actions ( $X_k$  or  $Y_k$ ), with each action being equally likely to be the privately preferred one. This can capture, for example, the fact that one action is much easier to implement or gives some career benefits. These private preferences are privately revealed (to the player executing the project) at the same time as the beliefs, i.e., in stage 1b, and are consumed (sunk) at the time of the action, i.e., in stage 2b.

To keep the focus on the original model, I will assume that these private benefits are never so large that they change the optimal action. A (strong) sufficient condition, which I will assume, is that  $b < (1 - p)(1 - 2\theta)(2\underline{v} - 1) - p\theta\underline{v}$ . The following proposition, then, states that the results of Proposition 1 continue to hold as long as the private benefits are not too large.

**PROPOSITION 2:** *As long as  $b < p \min(B + \underline{v} - \underline{u}, B)$ , the results of Proposition 1 extend with one modification: the wage in the  $F_I$  equilibrium must be  $w_k^{-i} > b/p \geq 0$ .*

**PROOF:**

See Web Appendix.

In this case, the efficiency wage must be high enough to make the employee forgo her private benefits rather than risk losing the wage. The condition that  $b < p \min(B + \underline{\nu} - \underline{u}, B)$  ensures that the maximal efficiency wage is sufficient to make  $P_{-i}$  obey, despite his private preferences. When  $b > p \min(B + \underline{\nu} - \underline{u}, B)$ , the private benefits may start to affect the firm boundaries. The next subsection considers these issues for a more traditional moral hazard setting.

The most interesting insight here is how the efficiency wage dominates pay for performance when there is both open disagreement and private costs/benefits (since pay for performance suboptimally allocates residual income). Even a small probability of disagreement could be sufficient to get this result.

### B. A Model with Moral Hazard

This subsection studies the effect of adding traditional moral hazard to the model. This serves multiple purposes. First, it clarifies the mechanisms that drive the results. In particular, it makes very clear that the results do not depend on making the agent indifferent about the course of action, and that the results are robust to the introduction of moral hazard. Second, this analysis relates the model to the extensive economic literature on moral hazard. Finally, although that is not the focus of the analysis, the model also provides a potential explanation for the trade-off between motivation and cooperation. The latter is considered one of the fundamental trade-offs in organization design and in the choice between markets and hierarchies (Roberts 2004).

Consider a setting where the project outcome depends also on unobservable private effort by the players. In particular, as indicated in Figure 3, in stage 3c—*after* it has been revealed whether the project is a success or a failure—each player can exert effort at private cost  $C_e$ . A player's effort affects the payoff of the project that he executes if (and only if) the project is a success: upon effort, a success always gives a payoff  $B + 1$ , while without effort a successful project gives a payoff  $B + 1$  only with probability  $(1 - \theta_e)$ , and  $B$  otherwise. Another way to state this is that the probability of a high payoff now equals  $Q_k = [1 - \theta_e I_{ne}][(1 - \theta)d_k + \theta d_{-k}]$ , where  $I_{ne}$  is the indicator function that the player executing the project did not exert effort. Assume, for simplicity, that  $C_e < \theta_e$  so that effort is efficient upon a (known) success, and reparametrize the cost of effort as  $c_e = C_e/\theta_e$ .

To keep the analysis tractable, I will make two simplifying assumptions. The first is that both decisions now depend on the same state variable:  $S_1 = S_2 = S \in \{X, Y\}$ .<sup>12</sup> The second is a parametric assumption that replaces the assumptions on  $\underline{u}$  from the main model:

ASSUMPTION 1:  $B + [1 - \max(\theta_e, c_e)][(1 - \theta)\nu + \theta(1 - \nu)] > \underline{u} > (1 - \theta_e c_e)\nu$ .

The upper-bound condition ensures that it will never be optimal to terminate the project on the equilibrium path. The lower-bound condition both ensures that the manager can always commit to firing the employee and simplifies the analysis considerably. To simplify the analysis further (on some knife-edge cases that turn out to be suboptimal), I will assume that when indifferent (financially), players' preferences, in order of importance, are to (i) undertake the action that they believe is most likely to succeed, (ii) communicate less, and (iii) stay.

The following proposition then shows that there are again market and firm regimes, as in Proposition 1. A key difference with Proposition 1 is that an employee who is subject to authority will now sometimes get outcome-contingent pay to generate incentives for private effort. Such

<sup>12</sup> In Section I, the two state variables enabled differential confidence, which drove nonintegration. Since private effort now takes that role, there is no need to keep this feature, which complicates the analysis considerably.

1	2	3
<p>Contracting</p> <p>a. The players negotiate asset ownership, project allocation, and project contracts.</p> <p>b. The beliefs <math>\mu^i</math> get (privately) drawn.</p>	<p>Orders and decisions</p> <p>a. Each player <math>P_i</math> can send a cheap-talk message from <math>\{X, Y\}</math>.</p> <p>b. Each player <math>P(k)</math> publicly chooses his action from <math>\{X, Y\}</math>. With probability <math>p</math>, the outside options drop to <math>\underline{u}_k^i = 0</math>.</p> <p>c. Each participant in a project can terminate that project.</p>	<p>Execution and payoffs</p> <p>a. Assets are committed (so that <math>\underline{u}_k^i = 0</math> for sure). Success or failure is revealed.</p> <p>b. Each participant in a project can terminate that project.</p> <p>c. Each player decides whether to exert effort (a private cost <math>C_e</math>).</p> <p>d. The contract terms <math>(w_k^i, \alpha_k^i)</math> get executed.</p>

FIGURE 3. TIMELINE OF MODEL WITH MORAL HAZARD

outcome-contingent pay must then be balanced by higher efficiency wages to keep him from disobeying.

**PROPOSITION 3:** *The parameters space is partitioned into three sets, denoted  $F_{NI}$ ,  $F_{I,fe}$ , and  $F_{I,pe}$ :*

- (i) *In  $F_{NI}$ —the nonintegrated (market) set—the subgame equilibrium starting in 2a is “no authority”; the only ownership structure that is part of an equilibrium for all parameter values has each player own one asset, say  $P_1$  owns  $a_1$  and  $P_2$  owns  $a_2$ ; each player then executes the project for which he owns the asset and gets the residual income of that project ( $\alpha_1^1 = \alpha_2^2 = 1$ ); both players exert effort.*
- (ii) *In  $F_{I,fe}$  and  $F_{I,pe}$ —the integrated (firm) sets—the subgame equilibrium starting in 2a is “authority by  $P_i$  over  $P_{-i}$ ”; the only ownership structure that is part of an equilibrium for all parameter values has  $P_i$  (the player with authority) own both assets.*
  - (a) *In  $F_{I,fe}$ —the integrated partition with “full effort”—the contract sets  $\alpha_1^i = 1$  and  $\alpha_2^i = 1 - c_e$  and both players exert effort.*
  - (b) *In  $F_{I,pe}$ —the integrated partition with “partial effort”—the player with authority gets all residual income of both projects ( $\alpha_1^1 = \alpha_2^1 = 1$ ) and only that player exerts effort.*

**PROOF:**

See Web Appendix.

In this model, it is the need to elicit effort that drives nonintegration. It does so in two ways. First, pay-for-performance incentives tempt the employee to disobey orders and this has to be counterbalanced by a higher efficiency wage. When the necessary efficiency wage becomes too high, it becomes optimal to sacrifice either effort or obedience. Second, shifting residual income to the employee also favors shifting control to the employee. At some point, it actually becomes optimal to let the employee go as an independent contractor, i.e., let him make his own decisions and bear the financial consequences of these decisions, and thus let him own the asset. This is

captured graphically in Figure 4, which also captures the trade-off between effort and cooperation. In particular, getting both effort and cooperation is optimal only when both are very important. When either or both are less important, it will be optimal to trade off one against the other.

The reason Proposition 3 is formulated in terms of “the only ownership structure that is part of an equilibrium for all parameter values” is that the outside options  $\underline{u}$  are now allowed to be very small, and assets with a very small opportunity cost are sometimes inconsequential.

Overall, this subsection thus shows how the results extend to a context with moral hazard, and how such moral hazard may become a force for nonintegration, inducing a trade-off between effort and cooperation. I will now discuss some other aspects of this theory of the firm.

### C. Discussion

*The Role of the Differing Priors Assumption.*—A first important issue is the role of differing priors in the results. In other words, which results depend on the differing priors assumption and which results extend beyond?

The direct effects of asset ownership on interpersonal authority—how the outside options affect the incentives to obey and to fire—do not depend on the differing priors assumption. They seem to extend, for example, to settings where the agency conflict is caused by effort and private benefits or to settings where the “contract” is a relational contract in a repeated-game setting (although the current setting does not need to invoke a repeated game mechanism to obtain authority).

Most of the other mechanisms in the paper, on the other hand, seem to depend to some extent on the assumption of differing priors or open disagreement. First, while the general principle of the colocation of income and control has been derived on other grounds (Aghion and Patrick Bolton 1992; Mathias Dewatripont and Tirole 1994), the specific mechanism in this paper seems to depend on differing priors (Van den Steen, forthcoming). In particular, under a common prior, players have on average identical valuations for the residual income—independent of who has control—and while players with more income care more about decisions being correct, this often means they prefer others to make the decisions (when the others are better informed). Second, the link between low-powered incentives and interpersonal authority also depends on differing priors. In particular, allocating residual income to the agent aligns his objectives with those of the principal in a common priors setting, so that high-powered incentives will not cause disobedience, but rather the opposite. Finally, while the link between asset ownership and the allocation of residual income does not depend on differing priors, its implication that assets should be allocated to players with control is caused by differing valuations induced by differing priors. For many of these results, however, a small probability of disagreement actually seems sufficient.

The part of the theory that depends on disagreement is most relevant in settings where fundamental disagreement is most likely: industries that face important strategic decisions under uncertainty, emerging industries with no established business models, and industries that undergo a fast or radical transformation. In the extreme, that part of the theory would not apply to firms in stable and mature industries with well-understood business models and that have no important uncertainty on the horizon, i.e., firms for which there is unanimity regarding which strategy and tactics will lead to success.

*Characterization of Relevant Assets.*—A second important question is whether this theory works for all assets or, if not, how to characterize the relevant assets. Probably the best characterization—and one that distinguishes this from the property rights theory (Grossman and Hart 1986; Hart and Moore 1990; Hart 1995)—is that the theory works for any asset that causes an opportunity cost when used in the project, and that assets with a larger opportunity cost are

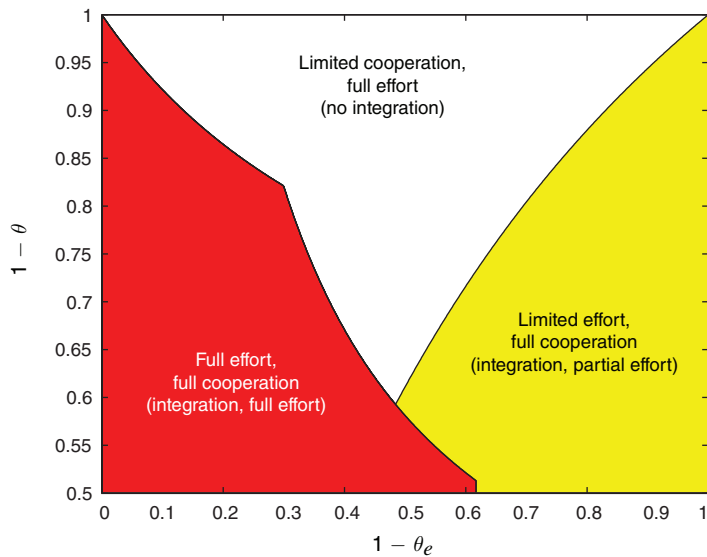


FIGURE 4. THE TRADEOFF BETWEEN EFFORT AND COOPERATION

Note:  $\nu = 0.75$ ,  $p = 0.1$ ,  $B = \underline{u} = 1$ .

more relevant. In particular, asset ownership in this theory determines the difference in utility between executing the project and taking the outside option, which is exactly the opportunity cost of using the asset. In the property rights theory, on the other hand, the relevant characteristic is not the level but the *slope* with respect to *investments* in specific human capital. As pointed out by Michael Whinston (2003), there is no obvious relationship between the level and that slope of the outside options. Moreover, unlike the property rights theory, the results of this paper are not affected by whether extra assets can be acquired on the market: for each project, the only relevant asset is the one that incurs an opportunity cost, i.e., the one that is really used in the project. The role of assets is thus very different here than in the property rights theory.

Closely related is the observation that—if assets could be freely traded—the relevant time of ownership is when the opportunity cost is realized, since that determines who incurs the cost. Therefore, the possibility of trading assets after the project is terminated does not affect the results.

Finally, by considering human capital as productive but inalienable assets that incur an opportunity cost, the theory also has implications for the effect of human capital on contracts and firms. First, important general human capital raises the outside options and may force the firm to give the human-capital providers a stake in the firm's residual income. Second, both general and specific human capital make firing employees more costly. It follows that both general and specific human capital of the employees reduce the effective interpersonal authority of the manager, while general human capital also increases the likelihood of self-employment and partnerships.

*Knight, Coase, and Firm Boundaries.*—The theory in this paper also suggests a new perspective on the Knightian versus Coasian views of the firm. To develop this perspective, let me start by arguing how the theory in this paper is closely related to Knight's (1921) theory of entrepreneurship. The key here is the observation that differing priors fit quite well Knight's notion of "uncertainty" (as

opposed to “risk”).<sup>13</sup> In particular, Knight suggested designating risk and uncertainty by, respectively, “the terms ‘objective’ and ‘subjective’ probability” (p. 233). He then argued that a person’s belief about an uncertain event may look like an objective probability, but that was a “confusion”: “[Instead] we do estimate...the dependability of our opinions or estimates, and such an estimate has the same *form* as a probability judgement...” (p. 231). Since a person’s prior is his personal subjective estimate of some (currently) unknowable probability, this fits closely with this paper: a person has an opinion that, say,  $X$  is the most likely state and  $\nu$  is his estimate of the dependability of that opinion, i.e., his “confidence” in the terminology of Knight (and of this paper).

According to Knight, then, people differ (with respect to uncertainty) in the confidence in their judgment and “the most fundamental change of all in the form of organization” is the “system under which the confident... ‘insure’ the doubtful... by guaranteeing to the latter a specified income in return for an assignment of the actual results” (p. 269). While this is often interpreted as risk aversion, Knight’s “confidence” actually refers to the belief that your judgement is correct, i.e.,  $\nu$ . This description *does* fit the mechanism in this paper: the principal tells the agent what to do and gives the agent a fixed wage in exchange, i.e., he *insures* the agent against the principal’s mistakes; and it is the more confident player who becomes the boss.

From this interpretation of Knight, the essence of a firm is that it allows the confident entrepreneur to make all decisions under uncertainty and then bear all the financial consequences of these decisions, i.e., absorb the uncertainty. The driving force in the existence of a firm is, according to Knight, the difference in confidence, which was also one of the driving forces in this model. This is quite different from the Coasian view, where it is the cost of transactions, i.e., the cost of contracting on decisions, that drives firm boundaries. To show that these do indeed differ, I will now apply the results of Van den Steen (2008)—which studies how differing priors affect the results of Coase (1960)—to this setting. In particular, I will show that even if the  $D_k$  were perfectly contractible, firm boundaries—as captured here in terms of asset ownership and allocation of residual income—still matter. In fact, a *lower* cost of contracting will sometimes lead to *more* firm rather than less firm. It thus shows that a firm is more than a second-best substitute for contracts on actions when the latter are difficult or costly.

For the formal analysis, consider the model with moral hazard, assume that the beliefs are publicly drawn in stage 1c, and assume that players can write an enforceable contract on the decisions  $D_k$  in stage 2a (using Nash bargaining). The rest of the game remains unchanged. Since the actions are now contractible, the definition of “authority by  $P_1$ ” also does not require any more that  $P_i$  sends a cheap talk message or that  $P_i$  quits in 2c if  $P_{-i}$  disobeys. In particular, the messages and  $P_i$  quitting when  $P_{-i}$  disobeys are not part of the definition anymore. With that change, the following proposition says that there are still firm and market regimes, but making actions contractible does change firm boundaries.

**PROPOSITION 4:** *With the decisions contractible, the result of Proposition 3 extends, except for the mapping of the parameter space into the partition elements: parts of  $F_{NI}$  and  $F_{I,pe}$  shift to  $F_{I,fe}$ .*

**PROOF:**

See Web Appendix.

The result is nicely captured in Figure 5: the left graph reprises Figure 4, while the right graph represents the different regimes under identical parameters, but now with actions contractible.

<sup>13</sup> While Knight did not provide a formal definition, people have tried to capture his idea of risk as “randomness with knowable probabilities” and of uncertainty as “randomness with unknowable probabilities.”

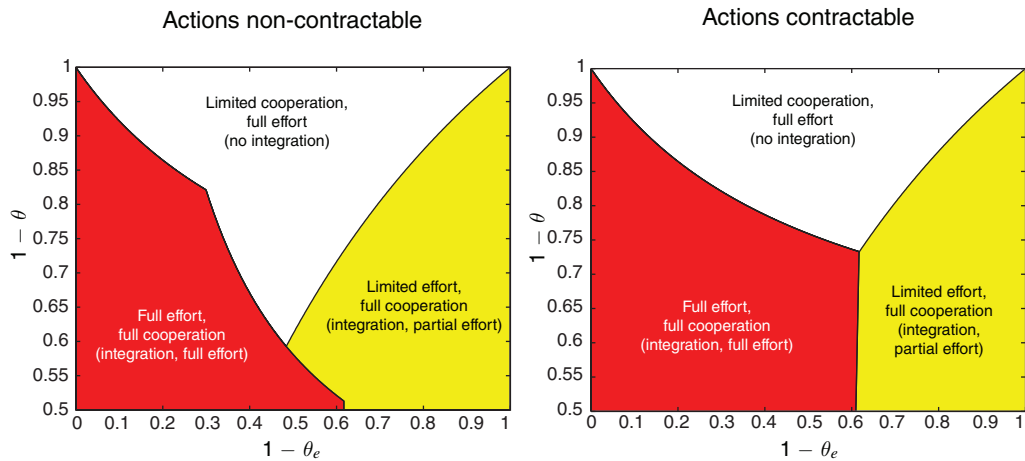


FIGURE 5. THE EFFECT OF CONTRACTIBILITY

Note:  $\nu = 0.75$ ,  $p = 0.1$ ,  $B = \underline{u} = 1$ .

This makes essentially two points. First, integration can be strictly optimal even when there is perfect and frictionless ex post bargaining over the decisions (and even absent incontractible investments that may cause hold-up). The intuition is that, due to their diverging beliefs, the two managers cannot agree on how to cooperate and—from each firm’s perspective—taking the wrong action is more costly than the joint gain. Integration effectively eliminates disagreement, since all projects are evaluated by the same person. This result is in marked contrast to the property rights theory of the firm (Grossman and Hart 1986; Hart and Moore 1990; Hart 1995)—where integration matters only in the presence of incontractible investments—and to Hart and Holmström (2002) or Baker, Gibbons and Murphy (2006), where ex post incontractibility is really necessary. A question that sometimes comes up in this context is why profit sharing cannot resolve the conflict: if both managers had a stake in both projects, would they not cooperate and therefore make integration unnecessary? The answer is that profit sharing does not align actions when people have different expectations over how actions pay off. It does, however, introduce inefficiencies by allocating residual income to players who value it less than others.

Second, contractibility *does* affect firm boundaries here, but not in the way that is often assumed. In particular, better contractibility leads her to more, rather than less, integration. The reason is that making actions contractible improves, in equilibrium, contracting *within* the firm: contracts allow the firm to function better. This analysis provides a new perspective on the Coasian versus Knightian view of the firm. In the Coasian view, the firm with its authority is a substitute for market contracting when the latter is too difficult or too expensive. In the Knightian view, the firm is a mechanism to allocate control and residual income under uncertainty to the most confident person (who is best at absorbing uncertainty). Lower contracting costs should always lead to less integration in the Coasian view, but not necessarily in the Knightian view.

*Disagreement, Break-up, and Firm Boundaries.*—This theory was partially motivated by personal observations of a real acquisition decision, which illustrates some parts of the theory but also raises some other issues. This particular acquisition was a horizontal integration: the focal firm wanted to fill out its product line and found another firm with a complementary product line.

The focal company was considering either an alliance or an outright acquisition. The main perceived risk of an alliance, relative to an acquisition, was the fear of future “strategic differences.” Not only would disagreement cause direct problems and losses, but both firms would have to make considerable relation-specific investments (integrating the product lines) that would be lost if disagreement proved unresolvable and caused a break-up of the alliance. These issues were instrumental in the firm’s decision to choose an acquisition, in order to have full control. Among other things, this obviously raises the question of why such disagreements could not be resolved through bargaining and contracting.

This paper explains why integration may indeed be optimal if the firm anticipates disagreement and a consequent break-up of the alliance. In fact, it shows that this may be optimal even in the absence of specific investments, and even if the future actions were perfectly contractible. Thus, this is a very different result from theories of hold-up, which fundamentally require specific and incontractible investments. The working paper (Van den Steen 2007c) did show, however, that specific investments exacerbate the issues in this paper. In particular, the anticipation of disagreement may prevent relation-specific investments, since the parties may fear that they will not reap the benefits of the investment. But, in this case, specific investments matter even when they are contractible. This is again in contrast to the hold-up literature, where hold-up matters only when it is caused by *noncontractible* ex ante investments. The reason why ex ante contractibility of the investment cannot solve the problem in this case is that the risk of disagreement and break-up remains. Overall, while the risk of break-up of an alliance due to disagreement has similarities to hold-up, it is generally a harder issue, since it causes direct costs and is more difficult to resolve through contracting. The prediction that the need for relation-specific investments leads to integration has found support in, for example, Kirk Monteverde and David J. Teece (1982).

*Empirical Implications.*—The theory has a number of empirical implications that can distinguish this theory from others. The following are some such implications. A high degree of uncertainty (for example in a new product category or when an industry faces fundamental changes) favors integration, especially when there are important decisions with decision externalities. Integration is also favored when coordination/cooperation is more important than effort/motivation, especially at high uncertainty. People who own important productive assets are less likely to be subject to authority. As a further consequence, important general and specific human capital both hinder interpersonal authority. General human capital further favors nonintegration and partnerships. Relation-specific investments will be higher within firms than between, and this relationship is stronger when there is more fundamental uncertainty or risk of disagreement around the activities related to the investments.

### III. Conclusion

This paper presents a theory of the firm in which asset ownership by a firm and low-powered incentive contracts for its employees serve the purpose of giving the firm’s manager interpersonal authority over the employees. In doing so, the theory also provides micro-foundations for the idea that bringing the project inside the firm gives the manager interpersonal authority over the employees working on the project and thus complements theories of rent-seeking and rent-seeking-style hold-up (Benjamin Klein, Robert Crawford, and Alchian 1978; Williamson 1985) or theories of adaptation (Simon 1951; Williamson 1975). I also argue that Knight’s view of entrepreneurship can be interpreted along the lines of this theory of the firm.

The firm boundaries in this paper were driven both by differential confidence on particular decisions and by the trade-off between effort and cooperation. By applying Van den Steen

(2008), I also show that firm boundaries may matter even when actions are perfectly contractible (ex ante or ex post). This result makes economic sense of the notion—often expressed by managers—that they may prefer outright mergers and acquisitions because they feel they need full control over the other firm if they will depend on it.

An important, but more implicit, aspect of this theory is its underlying conception of the firm as a legal person. This perspective is further developed in Van den Steen (2007b), which shows formally how this definition allows extending the current theory to a setting with multiple shareholders. In that case, the firm *aggregates* ownership and contracts to let the manager act “as if” he is the (sole) owner of all assets and party to all contracts, thus combining the authority benefits of concentrated asset ownership and the financial benefits of dispersed ownership. This “personal” theory of the firm also throws a new light on the discussions of firm boundaries, since firm integration is defined here as one legal person versus two, rather than a shift in boundaries, so that a clear definition of firm boundaries loses some of its importance.

The current paper also raises some further research questions. One interesting issue is that of authority between firms. Contrasting authority between firms with authority within firms should improve our understanding of firm boundaries, especially in relation to other arrangements such as alliances. This also raises the issue of other ownership structures, such as partnerships or consumer cooperatives. Another direction is the organization of international investments, where the issue of fundamental disagreement is well recognized (Stephen Hymer 1976). Finally, the paper also touches upon the central issue of firm boundaries. Peter Thompson and Steven Klepper (2006) provide interesting evidence on the role of strategic disagreement in the formation of new firms through spinoffs.<sup>14</sup> I believe that the current theory raises some interesting conjectures in that direction.

#### APPENDIX A: PROOFS

##### Proof of Proposition 1:

I start with a partial backward induction. Consider stage 3b. Upon a success, no participant quits in (any) equilibrium (given  $\alpha_k^i \geq 0$  and Nash bargaining). Upon a failure, a participant  $P_i$  quits  $R_k$  if and only if  $w_k^i < 0$ . Since  $w_k^1 + w_k^2 = B$ , it follows that if some  $w_k^i \notin [0, B]$ , then  $R_k$  is terminated upon failure, so that the  $w_k^i$  are paid only upon success. This case is equivalent to one where the payoff from  $R_k$  equals zero upon failure and  $1 + B$  upon success. It thus suffices to study the case where all  $w_k^i \in [0, B]$  and no participant quits in 3b, and then to show that getting  $B$  upon failure (and  $1 + B$  upon success) dominates getting zero upon failure (and  $1 + B$  upon success). I now assume that  $w_k^i \in [0, B]$  and no participant quits in 3b.

In stage 2c, given that “when indifferent the participant stays,” a participant  $P_i$  will quit  $R_k$  if and only if  $\underline{u}_k^i > w_k^i + \alpha_k^i [(1 - \theta)E^i(d_k) + \theta E^i(d_{-k})]$ , where  $E^i$  denote  $P_i$ 's expectation. This implies that

- (i) No participant will quit if all  $\underline{u}_k^i = 0$ ; and
- (ii) Only the participant who owns  $a_k$  will ever quit  $R_k$ . Furthermore,  $P_i$ 's decision to quit depends on the game history only through the values of  $\underline{u}_k^i$ ,  $d_k$ , and  $d_{-k}$ , and  $P_i$  is more likely to quit when  $D_m \neq Z_m^i$  than when  $D_m = Z_m^i$ . It also follows that a change in one of the actions cannot make  $P_j$  simultaneously strictly more likely to quit  $R_m$  and strictly less likely to quit  $R_{-m}$ .

<sup>14</sup> Although not stressed in the paper, they assume differing priors with regard to the precision of the information.

Consider next the decision in period 2b. Let  $I_{q,k}$  (with  $q$  referring to “quit”) be the indicator function that at least one player tries to quit  $R_k$  in 2c (which may depend on the  $D_m$ ). Player  $P_i$ 's expected utility equals  $U_i = E^i[\sum_k p I_{q,k} \underline{u}_k^i + (1 - p I_{q,k})[w_k^i + \alpha_k^i((1 - \theta)d_k + \theta d_{-k})]]$ . Let  $P_i = P(1)$ . Since 2b is a simultaneous-move game,  $P_{-i}$ 's decision for  $D_2$  is fixed from  $P_i$ 's perspective, so that only  $I_{q,m}$  and  $d_1$  can depend on  $D_1$ . Furthermore, if—given his information at the start of 2b— $P_i$  believes that the likelihood that  $P_{-i}$  quits either project is independent of his choice, then it is optimal for  $P_i$  to choose  $Z_1^i$ . For later reference, note that if  $P_2$  executes  $R_k$  (and never quits either  $R_m$ ) and  $P_1$  is committed to quitting  $R_k$  if and only if  $P_2$  disobeys (but never quits  $R_{-k}$ ), then  $P_2$  will obey upon disagreement if and only if  $w_k^2 + \alpha_k^2((1 - \theta)(1 - \nu_k^2) + \theta E^2(d_{-k})) + w_{-k}^2 + \alpha_{-k}^2((1 - \theta)E^2(d_{-k}) + \theta(1 - \nu_k^2)) > p\underline{u}_k^2 + (1 - p)[w_k^2 + \alpha_k^2((1 - \theta)\nu_k^2 + \theta E^2(d_{-k}))] + w_{-k}^2 + \alpha_{-k}^2((1 - \theta)E^2(d_{-k}) + \theta\nu_k^2)$  or if and only if  $w_k^2 > \underline{u}_k^2 + \alpha_k^2(((1 - p)/p)(1 - \theta)(2\nu_k^2 - 1) - (1 - \nu_k^2)) + \alpha_{-k}^2(1/p)\theta(2\nu_k^2 - 1)$ .

This concludes the backward induction. I will now derive particular equilibria. Rename, without loss of generality, the  $P_i$  and  $R_k$  so that  $\nu_1^1 + \nu_2^1 \geq \nu_1^2 + \nu_2^2$  and  $\nu_1^1 \geq \nu_2^1$ .

I will first argue that, for the subgame following the negotiation, there always exists a subgame equilibrium (denoted as  $F_I$ ) with  $(D_1, D_2) = (Z_1^1, Z_2^1)$ ,  $\alpha_1^1 = \alpha_2^1 = 1$ , and no player quits on the equilibrium path, and that, moreover, the elements of  $F_I$  in the proposition are necessary to optimally implement such equilibrium. This subgame equilibrium can always be implemented as follows:  $P_1$  owns both assets,  $P_1$  executes  $R_2$ , the contract sets  $w_1^1 \in [0, B] \cap [\underline{u} - (1 - \theta)\nu_1^1 - \theta\nu_2^1, \underline{u} - (1 - \theta)(1 - \nu_1^1) - \theta\nu_2^1]$  and  $w_2^1 \in [0, B] \cap [\underline{u} - (1 - \theta)\nu_2^1 - \theta\nu_1^1, B]$ , which are both always non-empty (given that  $\nu_1^1 \geq \nu_2^1$  and  $(1 - \theta)(1 - \bar{\nu}) + \theta\bar{\nu} \leq (1 - \theta)(1 - \underline{\nu}) + \theta\underline{\nu}$ ). In this case,  $P_2$  will never quit (since he owns no assets and  $w_1^2 > 0$  and  $w_2^2 \geq 0$ )—which immediately implies that  $P_1$  will always choose  $Z_2^1$ —while  $P_1$  will not quit when the actions are  $(Z_1^1, Z_2^1)$  (since  $w_1^1 + (1 - \theta)\nu_1^1 + \theta\nu_2^1 \geq \underline{u}$  and  $w_2^1 + (1 - \theta)\nu_2^1 + \theta\nu_1^1 \geq \underline{u}$ ) but will quit at least  $R_1$  when the actions are  $(\bar{Z}_1^1, \bar{Z}_2^1)$  (since  $w_1^1 + (1 - \theta)(1 - \nu_1^1) + \theta\nu_2^1 < \underline{u}$ ). The latter thus commits  $P_1$  to quit if  $P_2$  does not do as  $P_1$  wants. It follows that when  $Z_1^2 \neq Z_1^1$ , since  $P_2$  prefers the  $w_1^2 + w_2^2 > 0$  from choosing  $Z_1^1$  over either  $pw_1^2 + w_2^2 < w_1^2 + w_2^2$  (if  $P_1$  tries to quit  $R_1$ ) or  $pw_1^2 + pw_2^2 < w_1^2 + w_2^2$  (if  $P_1$  tries to quit both) from choosing  $Z_1^2$ ,  $P_2$  obeys (if he knows  $P_1$ 's belief). The equilibrium in the message game is then indeed that  $P_1$  (and only  $P_1$ ) tells  $P_2$  what to do. For most elements of the proposition, it is immediate that they are necessarily part of any optimal  $F_I$  equilibrium. For asset ownership, there are three arguments. First, if  $P_1$  does not own  $a_k$ , then he cannot commit to quit  $R_k$  upon disobedience so that he cannot use  $R_k$  to enforce obedience. Second, if  $P_2$  were to own  $a_k$ , then  $P_2$  would actually prefer the project to end as long as the asset's value is still  $\underline{u}$  (since his outside option  $\underline{u} > w_k^2$  for any  $w_k^2 \in [0, B]$ ) so that project termination is no threat—on the contrary—and  $P_2$  has no reason to obey. Third, and another way to formulate the second argument, if  $P_2$  owns  $a_k$  then his expected revenue from project  $R_k$  must be at least  $\underline{u}$ , which is impossible when  $\underline{u} > B$  and  $\alpha_k^2 = 0$ . The joint utility from  $F_I$  equals  $U_I = 2B + \nu_1^1 + \nu_2^1$ .

Let  $\nu$  denote  $(\nu_1^1, \nu_2^1, \nu_1^2, \nu_2^2)$  and define  $\hat{\theta} = (\bar{\nu} - \underline{\nu})/(2\bar{\nu} - 1)$  when  $\nu = (\bar{\nu}, \underline{\nu}, \underline{\nu}, \bar{\nu})$ , and  $\hat{\theta} = 0$  otherwise. I will next argue that (i) when  $\theta \leq \hat{\theta}$  (which requires that  $\nu = (\bar{\nu}, \underline{\nu}, \underline{\nu}, \bar{\nu})$ ) since  $\theta \in (0, 0.5)$  there always exists a subgame equilibrium (denoted as  $F_{NI}$ ) with  $(D_1, D_2) = (Z_1^1, Z_2^2)$ ,  $\alpha_1^1 = \alpha_2^2 = 1$ , and no player quitting on the equilibrium path; (ii) any  $F_{NI}$  equilibrium is dominated by the  $F_I$  equilibrium if and only if  $\theta > \hat{\theta}$ ; and (iii) any such equilibrium necessarily has the elements of  $F_{NI}$  in the proposition. To see that it always exists, consider the following:  $P_1$  owns  $a_1$  and  $P_2$  owns  $a_2$ , each player executes the project for which he owns the asset, and  $w_1^1 = w_2^2 \in [0, B] \cap [\underline{u} - (1 - \theta)\bar{\nu} - \theta(1 - \underline{\nu}), B]$  which is nonempty (since  $\theta \leq \hat{\theta}$  implies  $(1 - \theta)\bar{\nu} + \theta(1 - \bar{\nu}) \geq \underline{\nu}$  and thus  $B + (1 - \theta)\bar{\nu} + \theta(1 - \underline{\nu}) > B + \underline{\nu} > \underline{u}$ ). To see that this indeed implements that outcome, note that with these choice for  $w_k^i$  and  $\alpha_k^i$  neither player quits in 2c as long as he himself chose the action that he considered optimal (since  $w_2^1, w_1^2 \geq 0$  and  $w_1^1 + (1 - \theta)\bar{\nu} + \theta(1 - \underline{\nu}) \geq \underline{u}$ ). That further implies that each player will simply choose the action that

he considers most likely to succeed, and thus that there are no gains from sending messages in period 2a. For most of the elements of the proposition, it is again obvious that they are necessarily part of any optimal  $F_{NI}$  equilibrium. With regard to assets, each player must own the asset of his project since otherwise (with  $w_k^i \leq B$  and  $\alpha_1^i = \alpha_2^i = 1$ ) the asset owner would always want to quit the project. Another way to state this is that the player owning asset  $a_k$  must get at least some residual income  $\alpha_k^i$  (since  $B < \underline{u}$  and  $w_k^i \in [0, B]$ ) and that is only possible if  $P_1$  and  $P_2$  own, respectively,  $a_1$  and  $a_2$ . The joint utility for any equilibrium of this type equals  $U_{NI} = 2B + (1 - \theta)\nu_1^1 + (1 - \theta)\nu_2^2 + \theta$ . To determine when  $F_I$  dominates  $F_{NI}$ , note first that the naming of players and projects implies that  $U_{NI} \geq U_I$  requires  $\nu_2^2 > \nu_1^1$ , which then requires (by  $\nu_1^1 + \nu_2^2 \geq \nu_1^1 + \nu_2^2$ ) that  $\nu_1^1 > \nu_2^2$ , so that, indeed,  $(\nu_1^1, \nu_2^1, \nu_1^2, \nu_2^2) = (\bar{\nu}, \underline{\nu}, \underline{\nu}, \bar{\nu})$ . With all  $\nu_k^i$  fixed, it is now a matter of algebra to see that  $U_I > U_{NI}$  if and only if  $\theta > \hat{\theta}$ . This also implies that the equilibrium shifts from  $F_{NI}$  to  $F_I$  as  $\theta$  increases.

It now remains to be shown if these two equilibria dominate any others. Let  $S = (Z_1^1, Z_2^1, Z_2^1, Z_2^2)$  be the state in terms of the players' beliefs, let  $I_{q,k}(S)$  indicate whether one or both players try to quit  $R_k$  in stage 2c, and let  $\beta_k^i(S)$  indicate whether  $D_k = Z_k^i$ . The players' joint utility then equals

$$U = \sum_{S,k} \frac{1}{16} \left\{ pI_{q,k}(S)\underline{u} + (1 - pI_{q,k}(S)) \left[ B + \sum_i \alpha_k^i \left( \frac{1}{2} + (1 - \theta)(2\beta_k^i(S) - 1)(\nu_k^i - \frac{1}{2}) + \theta(2\beta_{-k}^i(S) - 1)(\nu_{-k}^i - \frac{1}{2}) \right) \right] \right\}.$$

I look for the optimum, disregarding feasibility. Since  $U$  strictly increases in the  $\beta_k^i(S)$ ,  $\beta_k^i(S) = 1$  when  $Z_k^i = Z_k^{-i}$ , and  $\beta_k^i(S) = 1 - \beta_k^{-i}(S)$  otherwise. The bilinear nature in terms of the  $\alpha_k^i$  and  $\beta_k^i(S)$  (with each  $\alpha_k^i$  interacting with a  $\beta_k^i(S)$  and vice versa) implies that  $\alpha_k^i, \beta_k^i(S) \in \{0, 1\}$ .

Consider, first, the case where  $\alpha_1^i = \alpha_2^i = 1$  for some  $i \in \{1, 2\}$ . In that case,  $U$  is maximized by also setting  $\beta_k^i(S) = 1, \forall k, \forall S$  (i.e., by following  $Z_k^i, \forall k$ ) and then by setting  $I_{q,k}(S) = 0$  given the assumption  $B + \underline{\nu} > \underline{u}_k$ . The joint utility then equals  $U = 2B + \nu_1^i + \nu_2^i$ . When  $P_i = P_1$ , this is exactly the  $F_I$  equilibrium. With  $P_i = P_2$ , this is weakly dominated by the  $F_I$  equilibrium (i.e., with  $P_i = P_1$ ) and strictly so unless  $\nu_1^1 = \nu_1^2$  and  $\nu_2^1 = \nu_2^2$ . This also implies that the more confident player will own the firm under integration.

Consider, next, the case where  $\alpha_1^i = \alpha_2^{-i} = 1$ . In that case,  $U$  is maximized either by  $\beta_k^1(S) = 1, \forall k, \forall S$ , or by  $\beta_1^i(S) = \beta_2^{-i}(S) = 1, \forall S$  (or by  $\beta_2^i(S) = \beta_1^{-i}(S) = 1, \forall S$ , but that is clearly dominated). The first case is strictly dominated by the  $F_I$  equilibrium. The second case is also strictly dominated by the  $F_I$  equilibrium unless  $(\nu_1^1, \nu_2^1, \nu_1^2, \nu_2^2) = (\bar{\nu}, \underline{\nu}, \underline{\nu}, \bar{\nu})$  and  $\theta \leq \hat{\theta}$ . In that case,  $B + (1 - \theta)\bar{\nu} + \theta(1 - \underline{\nu}) \geq \underline{u}$  so that at the optimum  $I_{q,k}(S) = 0$ . This is exactly the  $F_{NI}$  equilibrium. It is now, finally, straightforward to see that any equilibrium with  $w_k^i \notin [0, B]$  is indeed strictly dominated. This concludes the proof.

REFERENCES

Aghion, Philippe, and Patrick Bolton. 1992. "An Incomplete Contracts Approach to Financial Contracting." *Review of Economic Studies*, 59(3): 473-94.

Aghion, Philippe, and Jean Tirole. 1997. "Formal and Real Authority in Organizations." *Journal of Political Economy*, 105(1): 1-29.

Alchian, Armen A., and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization." *American Economic Review*, 62(5): 777-95.

Arrow, Kenneth J. 1964. "The Role of Securities in the Optimal Allocation of Risk-bearing." *Review of Economic Studies*, 31(2): 91-96.

- Arrow, Kenneth J. 1974. *The Limits of Organization*. New York: W. W. Norton & Co.
- Athey, Susan, and John Roberts. 2001. "Organizational Design: Decision Rights and Incentive Contracts." *American Economic Review*, 91(2): 200–05.
- Baker, George, Robert Gibbons, and Kevin J. Murphy. 2002. "Relational Contracts and the Theory of the Firm." *Quarterly Journal of Economics*, 117(1): 39–84.
- Baker, George, Robert Gibbons, and Kevin J. Murphy. 2006. "Contracting for Control." Unpublished.
- Boot, Arnoud W. A., Radhakrishnan Gopalan, and Anjan V. Thakor. 2006. "The Entrepreneur's Choice Between Private and Public Ownership." *Journal of Finance*, 61(2): 803–36.
- Brunnermeier, Markus K., and Jonathan A. Parker. 2005. "Optimal Expectations." *American Economic Review*, 95(4): 1092–1118.
- Calvo, Guillermo A., and Stanislaw Wellisz. 1979. "Hierarchy, Ability, and Income Distribution." *Journal of Political Economy*, 87(5): 991–1010.
- Chen, Joseph, Harrison Hong, and Jeremy C. Stein. 2002. "Breadth of Ownership and Stock Returns." *Journal of Financial Economics*, 66(2–3): 171–205.
- Coase, Ronald. 1937. "The Nature of the Firm." *Economica*, 4(16): 386–405.
- Coase, Ronald. 1960. "The Problem of Social Cost." *Journal of Law and Economics*, 3: 1–44.
- Dessein, Wouter, Luis Garicano, and Robert Gertner. 2006. "Organizing for Synergies." Center for Economic and Policy Research Discussion Paper 6019.
- Dewatripont, Mathias, and Jean Tirole. 1994. "A Theory of Debt and Equity: Diversity of Securities and Manager-Shareholder Congruence." *Quarterly Journal of Economics*, 109(4): 1027–54.
- Donaldson, Gordon, and Jay W. Lorsch. 1983. *Decision Making at the Top: The Shaping of Strategic Direction*. New York: Basic Books.
- Gibbons, Robert. 2005. "Four Formal(izable) Theories of the Firm?" *Journal of Economic Behavior and Organization*, 58(2): 200–45.
- Grossman, Sanford J., and Oliver D. Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, 94(4): 691–719.
- Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2006. "Does Culture Affect Economic Outcomes?" *Journal of Economic Perspectives*, 20(2): 23–48.
- Harrison, Michael, and David M. Kreps. 1978. "Speculative Investor Behavior in a Stock Market with Heterogeneous Expectations." *Quarterly Journal of Economics*, 92(2): 323–36.
- Hart, Oliver. 1995. *Firms, Contracts, and Financial Structure*. Oxford: Oxford University Press.
- Hart, Oliver, and Bengt Holmström. 2002. "A Theory of Firm Scope." MIT Department of Economics Working Paper 02–42.
- Hart, Oliver, and John Moore. 1990. "Property Rights and the Nature of the Firm." *Journal of Political Economy*, 98(6): 1119–58.
- Hart, Oliver, and John Moore. 2008. "Contracts as Reference Points." *Quarterly Journal of Economics*, 123(1): 1–48.
- Hermalin, Benjamin E. 1999. "The Firm as a Noneconomy: Some Comments on Holmström." *Journal of Law, Economics, and Organization*, 15(1): 103–06.
- Holmström, Bengt. 1999. "The Firm as a Subeconomy." *Journal of Law, Economics, and Organization*, 15(1): 74–102.
- Holmström, Bengt, and Paul Milgrom. 1994. "The Firm as an Incentive System." *American Economic Review*, 84(4): 972–91.
- Holmström, Bengt, and John Roberts. 1998. "The Boundaries of the Firm Revisited." *Journal of Economic Perspectives*, 12(4): 73–94.
- Hong, Harrison, and Jeremy C. Stein. 2007. "Disagreement and the Stock Market." *Journal of Economic Perspectives*, 21(2): 109–28.
- Hymer, Stephen. 1976. *The International Operations of National Firms: A Study of Direct Foreign Investment*. Cambridge, MA: MIT Press.
- Klein, Benjamin, Robert G. Crawford, and Armen A. Alchian. 1978. "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process." *Journal of Law and Economics*, 21(2): 297–326.
- Knight, Frank. 1921. *Risk, Uncertainty and Profit*. Chicago: University of Chicago Press.
- Landier, Augustin, and David Thesmar. 2009. "Financial Contracting with Optimistic Entrepreneurs." *Review of Financial Studies*, 22(1): 117–50.
- Levin, Jonathan. 2003. "Relational Incentive Contracts." *American Economic Review*, 93(3): 835–57.
- Levin, Jonathan, and Steven Tadelis. Forthcoming. "Contracting for Government Services: Theory and Evidence from U.S. Cities." *Journal of Industrial Economics*.

- Marino, Anthony, John Matsusaka, and Jan Zábajník.** Forthcoming. "Disobedience and Authority." *Journal of Law, Economics, and Organization*.
- Milgrom, Paul, and John Roberts.** 1988. "An Economic Approach to Influence Activities in Organizations." *American Journal of Sociology*, 94: 154–79.
- Monteverde, Kirk, and David J. Teece.** 1982. "Supplier Switching Costs and Vertical Integration in the Automobile Industry." *Bell Journal of Economics*, 13(1): 206–13.
- Morris, Stephen.** 1994. "Trade with Heterogeneous Prior Beliefs and Asymmetric Information." *Econometrica*, 62(6): 1327–47.
- Morris, Stephen.** 1995. "The Common Prior Assumption in Economic Theory." *Economics and Philosophy*, 11(2): 227–53.
- Morris, Stephen.** 1997. "Risk, Uncertainty and Hidden Information." *Theory and Decision*, 42(3): 235–69.
- Prendergast, Canice J.** 1995. "A Theory of Responsibility in Organizations." *Journal of Labor Economics*, 13(3): 387–400.
- Qian, Yingyi.** 1994. "Incentives and Loss of Control in an Optimal Hierarchy." *Review of Economic Studies*, 61(3): 527–44.
- Rajan, Raghuram G., and Luigi Zingales.** 1998. "Power in a Theory of the Firm." *Quarterly Journal of Economics*, 113(2): 387–432.
- Roberts, John.** 2004. *The Modern Firm: Organizational Design for Performance and Growth*. Oxford: Oxford University Press.
- Schein, Edgar H.** 1985. *Organizational Culture and Leadership*. San Francisco: Jossey-Bass Publishers.
- Scheinkman, Jose A., and Wei Xiong.** 2003. "Overconfidence and Speculative Bubbles." *Journal of Political Economy*, 111(6): 1183–1219.
- Simon, Herbert.** 1947. *Administrative Behavior*. New York: Free Press.
- Simon, Herbert.** 1951. "A Formal Theory of the Employment Relationship." *Econometrica*, 19(3): 293–305.
- Thompson, Peter, and Steven Klepper.** 2006. "Intra-Industry Spinoffs." Florida International University Department of Economics Working Paper 0605.
- Van den Steen, Eric.** 2004. "Rational Overoptimism (and Other Biases)." *American Economic Review*, 94(4): 1141–51.
- Van den Steen, Eric.** 2007a. "The Cost of Incentives under Disagreement: Too Motivated?" Unpublished.
- Van den Steen, Eric.** 2007b. "An Economic Theory of the Firm as a Legal Entity." Unpublished.
- Van den Steen, Eric.** 2007c. "Interpersonal Authority in a Theory of the Firm." MIT Sloan Research Paper 4667–07.
- Van den Steen, Eric.** 2008. "Disagreement versus the Coase Theorem." Unpublished.
- Van den Steen, Eric.** Forthcoming. "Disagreement and the Allocation of Control." *Journal of Law, Economics, and Organization*.
- Varian, Hal R.** 1989. "Differences of Opinion in Financial Markets." In *Financial Risk: Theory, Evidence and Implications*, ed. C. C. Stone, 3–37. Norwell, MA: Kluwer Academic.
- Wernerfelt, Birger.** 1997. "On the Nature and Scope of the Firm: An Adjustment-Cost Theory." *Journal of Business*, 70(4): 489–514.
- Wernerfelt, Birger.** 2002. "Why Should the Boss Own the Assets?" *Journal of Economics and Management Strategy*, 11(3): 473–85.
- Whinston, Michael D.** 2003. "On the Transaction Cost Determinants of Vertical Integration." *Journal of Law, Economics, and Organization*, 19(1): 1–23.
- Williamson, Oliver.** 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: Free Press.
- Williamson, Oliver.** 1985. *The Economic Institutions of Capitalism*. New York: Free Press.
- Wilson, Robert.** 1968. "The Theory of Syndicates." *Econometrica*, 36(1): 119–32.
- Yildiz, Muhamet.** 2003. "Bargaining without a Common Prior: An Immediate Agreement Theorem." *Econometrica*, 71(3): 793–811.