Interpersonal Authority in a Theory of the Firm

Eric Van den Steen*

February 8, 2007

Abstract

This paper proposes a theory of the firm in which a firm's centralized asset ownership and low-powered incentives are mechanisms to give a manager 'interpersonal authority' over employees, in a world with differing priors. The paper thus provides micro-foundations for the idea that bringing a project inside a firm gives the manager authority over that project, while – in the process – explaining concentrated asset ownership, low-powered incentives, and centralized authority as typical characteristics of firms. It also uses this theory to derive new results for firm boundaries.

I study 'interpersonal authority' (i.e., the ability of a manager to make subordinates obey her orders), as opposed to the 'decision authority' (i.e., her ability to make an impersonal decision) that has been more common in the literature. I derive interpersonal authority as an equilibrium phenomenon that arises through an efficiency-wage type contract between two originally symmetric players. Shifting asset ownership from an agent to a principal strengthens the principal's interpersonal authority through a change in the outside options of the efficiency-wage contract, while a flat wage makes the agent willing to obey orders that he disagrees with. Based on this, I show that one party should own all the assets for a project and that that owner should also be the project's residual claimant. This owner hires employees for the project under low-powered incentives (or a fixed-wage contract), while the employees accept orders from the owner. Different projects are often optimally owned by different people.

As an application of this theory to firm boundaries, I propose a new theory for integration: the risk of 'break-up'. In particular, I show that fundamental disagreement may cause two separate firms to go their own ways despite coordination being efficient from an outsider's perspective. Beyond the direct losses involved, the anticipation of such break-up may also prevent relation-specific investments. I show that a merger may be strictly optimal, *even when the relation-specific investments are ex-ante contractible and there is perfect ex-post Coasian bargaining about the decisions.* By unifying all interpersonal authority in the hands of one owner, the merger eliminates the risk of future fundamental disagreement.

1 Introduction

Interpersonal authority (i.e., the ability of a superior to tell her subordinates what to do) is a cornerstone of organization. It is most visibly expressed in the chain of command or in the hierarchy

*MIT-Sloan School of Management (evds@mit.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, Oliver Hart, David Scharfstein, Ravi Singh, and especially Bengt Holmstrom for many conversations that helped me structure my thinking on these issues. I finally thank the participants in the MIT organizational economics lunch, the Harvard-MIT organizational economics seminar, and the Stanford GSB-Economics seminar for their comments and suggestions. *The paper is currently in the process of being revised.*

of authority, which is the ranking of who can tell whom what to do. Being employed is so intertwined with a boss telling the employee what to do, that people are often said to 'become their own boss' when they strike out on their own. This raises the question how interpersonal authority and the firm relate.

The purpose of this paper is to develop a theory of the firm in which a firm's centralized asset ownership and low-powered incentives (or fixed wages) are mechanisms to give a manager interpersonal authority over employees, and to use that theory to derive implications for firm boundaries. I will say that one person has interpersonal authority over another person if: 1) the first person tells the second what to do, 2) this (literal) order makes the second person more likely to do what the order told him to do, and 3) the person obeying (sometimes) acts against his own beliefs or preferences. The third criterion distinguishes authority from simple advice. I thus use 'authority' in the sense of 'the power or right to give orders and enforce obedience' (Concise Oxford English Dictionary), which is essentially the definition given by Fayol (1916, 1949).

Note that I distinguish 'interpersonal authority' from 'decision authority' (i.e., the ability of a manager to make a particular decision). While some recent literature has studied decision authority (Marschak and Radner 1972, Milgrom and Roberts 1988, Prendergast 1995, Aghion and Tirole 1997, Baker, Gibbons, and Murphy 1999, Aghion, Dewatripont, and Rey 2004), my use of 'authority' is consistent with an earlier literature (Coase 1937, Simon 1951).[1]

The idea that authority plays a central role in the nature and function of a firm has a long and respectable tradition in economics. Knight (1921), Coase (1937), Simon (1951), Arrow (1974), and Williamson (1975) all interpreted or defined the firm as being about authority. Coase, for example, likens the firm to a 'master and servant' relationship. Alchian and Demsetz (1972), however, sharply criticized this view. In particular, they 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 to the Alchian-Demsetz critique, economists have looked for foundations other than authority to build a theory of the firm and define a firm's boundaries. The most influential approach has been the property rights theory of the firm (Grossman and Hart 1986, Hart and Moore 1990, Hart 1995). This theory defines a firm as a set of assets and makes predictions about who should own which assets. Ownership – in this theory – essentially provides bargaining chips to appropriate a larger part of the residual income, which in turn provides incentives to invest. Interpersonal authority plays no role in this property-rights theory.

In addition to responding to the Alchian-Demsetz critique, the property rights theory made two influential methodological contributions: its focus on asset ownership as a characteristic of firms and its insistence on holding the economic environment fixed, i.e., on not simply postulating changes from bringing a transaction inside the firm (Hart 1995, Gibbons 2005). This paper builds on these two methodological contributions of the property rights theory, but returns to the question *why (or whether) bringing an activity inside a firm would give the manager interpersonal authority over the people involved.*[2] In particular, I start from the following three observations: 1) firms hire people under low-powered incentive contracts to work on the firm's projects (Knight 1921, Simon

[1]Note also that these two concepts, while distinct, are closely related: the purpose of interpersonal authority is often to get decision authority, while decision authority often presumes interpersonal authority over the people implementing the decision. Since the literature has not always been explicit on the distinction, I will at times also refrain from making the distinction when discussing earlier literature.

[2]It is important to keep in mind, though, that firms are very complex institutions and simply cannot be understood from just one perspective to the exclusion of others. For an overview of this richness, see Holmstrom and Roberts (1998) or Gibbons (2005).

1951, Holmstrom and Milgrom 1991), 2) a firm's manager has interpersonal authority over these employees (Knight 1921, Coase 1937, Simon 1951, Arrow 1974, Williamson 1975), and 3) firms own the assets that are necessary for their activities (Hart 1995, Holmstrom 1999). I will argue that a firm's centralized asset ownership and fixed-wage contracts emerge endogenously as ways to give the firm's manager interpersonal authority over the firm's employees. Interpersonal authority and the need to give the manager such authority are themselves equilibrium outcomes of the model. Based on this, I show that the following bundle of practices is optimal:

- All assets that are necessary for one project should be owned by one party. The owner of the project's assets should also be its residual claimant.
- Other people working on the project should receive low-powered incentives or even fixed wages from the owner.
- In equilibrium, the owner tells these employees what to do, and employees obey the owner.

The paper thus provides micro-foundations for the idea that bringing an activity inside a firm gives the manager authority over that activity, while – in the process – explaining concentrated asset ownership and low-powered incentives as characteristics of a firm. I will also argue that this theory is closely related to Knight's (1921) theory of entrepreneurship or theory of the firm. Note further how this theory relates to earlier approaches and to the Alchian-Demsetz critique: instead of *assuming* that a firm conveys authority, it *explains* the firm as a mechanism to create such authority.

To derive these results, I study a setting in which a number of people can jointly undertake a project. A project is defined as a revenue stream that requires assets and that depends on decisions made by the participants (and potentially on private effort). A key issue is that the participants openly disagree – or have differing priors – on the right decisions. Decisions are not contractible but the project's outcome is. Moreover, each player can tell others – through cheap talk messages – what he believes others should do. Finally, cooperation is to a large degree at will: people can walk away from the project. This ability to walk away allows the parties to write efficiency-wage type contracts: by paying more than the market wage, the threat of firing can make one person obey the other. However, the ability to walk away also puts limits on the wages that can realistically be promised. The question is then who should own which assets and write what contracts with whom.

As hinted above, players will – in equilibrium – endogenously create interpersonal authority for one player over others by writing efficiency-wage type contracts (that also make it optimal for that one player to end their contract if the other disobeys). The cheap talk messages will play the role of (non-binding) 'orders'. The 'employee' obeys such non-binding orders because he realizes that the contract commits the owner to firing him if he disobeys and he rather obeys an order which he believes is wrong than to get fired.

The intuition behind the paper's key results is then based on three effects. First, moving asset ownership from the agent to the principal makes it more costly for the agent to get fired, since the agent has lower outside options. It also makes it easier for the principal to commit to firing a disobeying agent, since the principal keeps the assets. Both effects give the agent more reason to obey. Asset ownership by the firm thus strengthens the manager's interpersonal authority over the employees. Second, high-powered incentives give an agent a reason to follow his own beliefs when he disagrees with his principal, and thus lead to disobedience (Van den Steen 2005b). As a consequence, low-powered incentives for the employee also increase the manager's interpersonal authority. Finally, the need for interpersonal authority itself derives from the result that it is optimal – under differing priors – to concentrate all control and income of a project in one hand (Van den Steen 2006a).

It is important to stress here that interpersonal authority – despite being widely used – poses a real challenge for organizations: a boss's orders often get disobeyed. Such disobedience can come in many forms, such as feigned ignorance, forgetfulness, sloppy work, or even purposeful errors. In his seminal work on organizations, Barnard (1938) wrote, for example, that '(authority) is so ineffective that the violation of authority is accepted as a matter of course and its implications are not considered. It is surprising (...) how generally orders are disobeyed (...).' There is also a drill sergeant's saying that 'The army cannot make you do something, but it sure as hell can make you wish you had.' In the end, people have a free will and it is up to the organization to make them obey. It is for this reason that the need to induce employees to obey orders can indeed be a core determinant of organization design and performance.

As an application of this theory of the firm to the question of firm boundaries, I propose a new theory for integration: the risk of 'break-up'. In particular, I show how fundamental disagreement on the right course of action may cause an ex-post misalignment between the actions of two separate firms that – in order to prevent the disagreement and thus the misalignment – may make their integration ex-ante strictly optimal. The new, and probably surprising, element in this result is to show that, even though any outsider would consider it efficient for both firms to align their actions, the managers' disagreement may make them unable to agree on what action to coordinate on (*even* if there is perfect ex-post Coasian bargaining on the actions!) The effect of integration is to eliminate such disagreement by giving one manager interpersonal authority over all employees. A further result is that the anticipation of such fundamental disagreement and break-up may prevent relation-specific investments that require future coordination. Integration will increase specific investments and may then be strictly optimal *even when I allow perfect contractibility of the specific investments and perfect ex-post Coasian bargaining on the decisions.* This also captures the idea that a merger may often dominate an alliance because it provides more control.

The theory in this paper is formulated in terms of a manager-owner or entrepreneur-owner. While this approach has a long tradition – starting with Knight (1921) and Coase (1937) – it obviously raises the question whether the model can be used to study large firms, such as Ford or GE. As discussed in more detail in subsection 4.1, the formal definition of a firm in this paper is that of a firm as a legal person that can write contracts and own assets. Van den Steen (2006b) – which builds on this paper – shows how this definition makes it possible to apply this theory to firms with multiple shareholders. Since the current paper considers only manager-owned or entrepreneur-owned firms that do not change hands, the firm (as a legal person) and the owner (as a physical person) are interchangeable. There is thus no reason to distinguish them formally here. For transparency reasons, the whole model is therefore formulated in terms of managers-owners.

Contribution This paper makes three contributions. First, it shows how asset ownership plays a role in determining interpersonal authority, and thus in determining effective control. Second, building on this and other results, it shows how a firm's concentrated ownership and flat wages give the manager interpersonal authority over the firm's employees. The paper thus provides micro-foundations for the idea that bringing an activity inside a firm gives the manager authority over that activity, while – as part of that – explaining concentrated asset ownership and flat wages as typical characteristics of firms. Third, it introduces the risk of break-up as a reason for integration. It also shows how the anticipation of break-up – much like the anticipation of hold-up – may prevent relation-specific investments (even if the investments are contractible) and thus lead to integration.

All three contributions, however, are about one thing: how the firm is essentially a mechanism to give the manager interpersonal authority over the firm's employees.

Literature To relate this paper to the literature on the theory of the firm, it is useful to start from Gibbons's (2005) distinction between the 'control branch' and the 'contract branch' of the theory of the firm . The 'control branch' simply *asserts* that integration gives a manager authority and includes Knight (1921), Coase (1937), Simon (1951), and Williamson (1975) as some of the most well-known contributors. The 'contract branch' denies that integration changes anything and just uses the firm as a label for a set of contractual relationships. It includes Alchian and Demsetz (1972), the Grossman-Hart-Moore property rights theory, Holmstrom and Milgrom (1994), Rajan and Zingales (1998), and Levin and Tadelis (2004), among others. The difference between this paper and the control branch is that, instead of *assuming* or *asserting* that the firm gives the manager interpersonal authority, I formally derive the firm – with its centralized asset ownership and low-powered incentives – as a mechanism to generate interpersonal authority. In that sense, the paper would be part of the contract branch while providing micro-foundations for the control branch. The paper differs, however, from the existing literature in the contract branch by its focus on the question what it is about firms that gives a manager 'interpersonal authority' over employees or that makes employees obey their manager (i.e., the 'master and servant' relationship of Coase or the grocer versus employee issue of Alchian and Demsetz) and by conceptualizing the firm as a distinct entity, a legal person.[3]

The question how the need to generate interpersonal authority affects organizations was – to my knowledge – first studied in Van den Steen (2005b), which shows that pay-for-performance may hinder interpersonal authority, so that interpersonal authority will go together with low-powered incentives or even fixed wages. That paper also shows that people with strong beliefs and high intrinsic motivation will be more likely to become independent entrepreneurs. I discuss the link with the current paper below. In a recent related contribution, Marino, Matsusaka, and Zábojník (2006) study the reverse problem: how organizational and market characteristics – such as the agent's job market – affect obedience and thus the equilibrium allocation of control. They do not consider the role of assets or how this feeds into a theory of the firm.

The more specific idea that asset ownership may affect authority has been suggested before, but in a very different sense than the interpretation or formalization in this paper. Hart (1995) mentions the idea that asset ownership can convey authority, but his formalization (p.61) shows that he has something very different in mind than one person giving orders to another person and being obeyed. In particular, Hart shows that asset ownership makes others orient their specific investments towards the asset owner, which he then interprets as asset ownership conveying authority. Holmstrom (1999), in his theory of the firm as a subeconomy, is the first to suggest ideas that are closer to this paper. He argues that 'asset ownership conveys the CEO (...) the ability to restructure the incentives of those that accept to do business (in or with the firm)'. Ownership then conveys a form of decision authority, but not the interpersonal authority that is the focus of this paper. Hermalin (1999) pushes this a step further and argues informally that centralization of control rights may prove difficult without also centralizing asset ownership. In particular, he suggests that

[3] Two other recent contributions are Hart and Holmstrom (2002) and Hart and Moore (2006). The first suggests a theory of firm boundaries based on the fact that firm boundaries determine who is in control, and thus what private preferences and what beliefs that person brings to the table. The second suggests explanations for employment contracts and vertical integration based on the assumption that explicit contracts provide a reference point for people's feeling of entitlement. Neither paper deals, however, with the central question of this paper why an employee would obey his manager or would be more likely to do so within a firm.

if employees own assets, then they can force decisions by threatening holdup or exit. His argument, however, remains very different from the current paper. In particular, the argument in this paper is essentially the reverse: asset ownership by the firm makes the threat of firing the employee both more credible (since the firm retains the assets) and more powerful (since the employee loses the assets and thus has lower outside options), increasing the interpersonal authority of the manager over the employee. Wernerfelt (2002) presents an argument that is a reverse in a different sense: he shows that the person in control – the boss – should own the assets, since he can better internalize the effect of his decisions on the assets.[4]

One important feature of this paper is that it simultaneously derives the triple centralized asset ownership, low-powered incentives (or fixed wages), and interpersonal authority as characteristics of firms, and shows how they are closely related. Holmstrom and Milgrom (1994) long preceded this paper with a somewhat similar feature. In particular, they simultaneously explain – by combining monotone comparative statics (Milgrom and Roberts 1990) with multi-tasking (Holmstrom and Milgrom 1991) – the following triple: employees do not own assets, employees have low-powered incentives, and the firm can exclude employees from certain returns (such as the ability to take outside jobs). While excluding employees from certain returns can be done contractually, it can also be done by using authority. When taking that perspective, their paper also deals with the triple asset ownership, low-powered incentives, and authority. One key difference is what is meant with authority: their paper deals with the *use* of authority to forbid employees to receive income from outside activities, while this paper deals with the *origin* of interpersonal authority that is used to *directly* tell employees what to do and what not to do in a fairly general sense. Another important difference is in the role of assets. In Holmstrom and Milgrom (1994), it does not matter who owns the assets, as long as they are not owned by the employee, so that shifting assets from one firm to another does not matter, in contrast to the current paper. This latter distinction is important when it comes to discussing firm boundaries. A key insight of Holmstrom and Milgrom (1994) that has influenced this paper a lot is the observation that low-powered incentives in firms may not simply be an unfortunate consequence but the explicit purpose of transacting through a firm. Their paper is also the first to explicitly think about the firm in terms of a set of *complementary* practices.

The current paper relies heavily on three earlier papers on contracting under differing priors.[5] Since these papers are so intertwined with the current one, it is worth spelling out the contributions in more detail. The three earlier papers have a common origin[6] and deal essentially with the same broad issue: the interaction between disagreement and control in a multi-person project. The first of these papers, Van den Steen (2005b), was discussed above. It delivers the result that interpersonal authority will go together with low-powered incentives, which is a key element of the current paper. The second of these papers, Van den Steen (2006a), assumes that authority can be allocated contractually and asks how to optimally allocate authority in the absence of motivation and coordination issues. It shows, among other things, that authority over complementary decision should be co-located while authority over substitute decisions should be distributed, and that authority and residual income should be co-located when people have differing priors. The latter

[4]Wernerfelt (1997) uses a repeated-game argument to argue why a manager has authority over employees. This raises the question, however, how such repeated-game argument differs from repeated-game arguments in market relationships. Turnover in many firms lead to much shorter firm-employee relationships than the relation between, say, Microsoft and Intel.

[5]A complementary literature, though with different focus, is that on financial contracting under differing priors, such as Boot, Gopalan, and Thakor (2006) or Dittmar and Thakor (2006).

[6]All three paper are derived from one broad exploratory working paper 'Interpersonal Authority: A Differing Priors Perspective' (2004).

result is a key element in the current paper and its intuition will be discussed in more detail later. The third paper, Van den Steen (2006c), again assumes that authority can be allocated contractually, and shows that allocating authority induces a trade-off between motivation and coordination, which is a well-known issue in organization design. A key part of that paper's analysis is to show that differing priors often require the use of authority (instead of incentives) to achieve coordination. Section 6 relies in part on this mechanism to propose a new theory for firm boundaries. The contribution of the current paper is to put all these results in a theory-of-the-firm context (with the firm defined as a legal person), to add asset allocation as another lever, and to show that this all adds up to a theory of the firm and to new implications for firm boundaries.

The following section describes a simple version of the model (which takes the outside options as exogenous and limits the number of assets and players). Section 3 uses that model to derive the key results. Section 4 discusses some important foundations and interpretations of the theory, including the definition of the firm as a legal person and the relationship to Knight's theory of entrepreneurship. Section 5 endogenizes the outside options. Section 6 shows how this theory can explain why the risk of break-up – due to 'strategic differences' – may lead to changes in firm boundaries, while section 7 concludes.

2 The Model

The model in this paper captures a setting in which two people can start a project together if they have the necessary assets. The key issue is that the two participants may openly disagree on the right course of action. To resolve this conflict, the participants can try to structure their contract such that one participant effectively controls the other. In particular, one participant can – in the style of an efficiency wage – promise the other a high wage and then threaten to 'fire' the other (by ending the contract) if the latter 'disobeys'. Asset ownership then affects the participants' outside options when the project is ended prematurely (i.e., when someone gets fired) and thus the control that one player has over the other. In this section, I take the outside options as exogenous to make the analysis maximally transparent. Section 5 will endogenize the outside options by allowing each player to rematch and execute the project with some other player if the first project ends prematurely. (Section 5 will also allow for more players, more projects, and more assets.) The model includes (some) moral hazard and intrinsic motivation, mainly to show that the mechanisms also work in such setting.

Formally, consider two players (denoted P_1 and P_2) and two assets (denoted a_1 and a_2). The players can engage in a 'project', which is a revenue stream R that requires both players and both assets. As part of this project, each participant P_i has to make a decision $D_i \in \{X, Y\}$. One and only one of these choices is correct, as captured by the state variable $S \in \{X, Y\}$, which happens to be common to both decisions (although that is not necessary for the results).[7]

As depicted in figure 1, the project will be either a success or a failure. A failure always gives payoff $B > 0$, while a success gives payoff $B+1$ with probability η, and B otherwise. This probability η may depend on the players' effort, as discussed later.[8] The probability of success itself depends on which decisions are correct. In particular, let $d_i = I_{\{D_i=S\}}$ be the indicator that P_i's decision

[7]In particular, in section 6 the states will be decision-specific.

[8]This moral hazard is not at all necessary for the analysis. It is introduced merely to show that the model works in a setting with moral hazard. (In particular, I will also show the results for $\eta \equiv 1$ independent of any effort.) The same is true for the intrinsic motivation, i.e., private benefit from success, that I will discuss later.

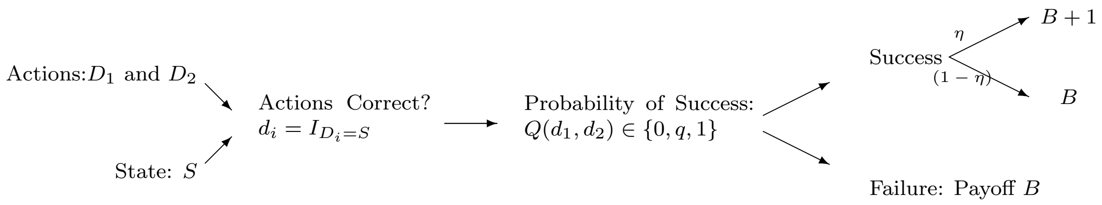


Figure 1: Actions and Payoffs

D_i is correct and let the probability of success be $Q(d_1, d_2) \in [0, 1]$, with Q symmetric and strictly increasing in d_1 and d_2; and with $Q(0, 0) = 0$, $Q(1, 1) = 1$, and $Q(0, 1) = Q(1, 0) = q \in (0, 1)$.

The state S is unknown, but each player P_i has a subjective belief μ_i that $S = X$. A key assumption is that (it is common knowledge that) players have differing priors, i.e., they can disagree in their beliefs about S even though neither player has private information about S. The fact that these are differing priors and there is no private information about S also implies that players will not update their beliefs when they notice that someone else has a different belief: they simply accept that people sometimes disagree.[9]

To keep the analysis simple, both players' beliefs will be independent draws from a commonly known binary distribution: for some $\nu \in (.5, 1)$, μ_i equals ν or $(1-\nu)$ with equal probability. This implies that i believes half the time that X is the best course of action, and half the time that Y is the best course of action. But i always has the same confidence (or strength of belief in what he believes is best) $\nu = \max(\mu_i, 1-\mu_i)$. Moreover, the two players disagree half the time. In the main analysis, both players have the same confidence ν in order to make clear that the results are not driven by asymmetries between the players. I will, however, also consider what happens when players may differ in their level of confidence.

The timing of the game is indicated in figure 2. First, the players negotiate a contract via symmetric Nash bargaining, with outside values as described later.[10] To discuss the contract, I

[9]For a more extensive discussion of differing priors, see Morris (1995) or Van den Steen (2001). Some papers by other authors that have used differing priors include Harrison and Kreps (1978), Morris (1994), Yildiz (2003), or Boot, Gopalan, and Thakor (2006). Differing priors do not contradict the economic paradigm: while rational agents should use Bayes' rule to update their prior with new information, nothing is said about those priors themselves, which are primitives of the model. In particular, absent any relevant information, agents have no rational basis to agree on a prior. Harsanyi (1968) observed that 'by the very nature of subjective probabilities, even if two individuals have exactly the same information and are at exactly the same high level of intelligence, they may very well assign different subjective probabilities to the very same events'. Alternatively, differing priors for a particular model may originate from noisy Bayesian updating from an earlier common prior (with players being only partially aware that they may make mistakes). In other words, the term 'prior' is not used in an absolute sense, but relative to the timing of the model. In this paper, I am agnostic about the exact source of disagreement. It is important to note, however, that differing priors is *not* the same as private information that cannot be communicated, which would lead to very different predictions in this model.

[10]We could also allow the outside options to be zero or some other fixed values. Logically, however, the outside value at this point should be at least as high as the one later in the game. The paper's setup is the simplest way to ensure this. There are also many variations on this bargaining procedure that would work, such as take-it-or-leave-it offers in a random or fixed sequence or asymmetric Nash bargaining.

1	2	3	4
Contracting	Orders and Decisions	Execution	Payoffs
a The players negotiate contract (v_i, α_i, F_i) (via Nash bargaining). b F is paid instantly. c The beliefs μ_i get (privately) drawn.	a Each player can send a cheap-talk message from $\{X, Y\}$. b Each player publicly chooses his action from $\{X, Y\}$. c With probability p, each player can end the contract (prior to stage 3), in which case each P_i gets his outside option $\underline{u}_i$.	a State is realized, assets are committed. b Each player can end the contract, in which case each P_i gets his outside option 0.	a The players choose simultaneously whether to exert effort. b The contract terms (v_i, α_i) get executed.

Figure 2: Time line of basic model

need to specify what is contractible and what is not. I will assume that the decisions D_i (in stage 2b) are not contractible, and neither the right to make these decisions. However, the project success *is* contractible. A contract will then consist of a fixed payment v_i, a share $\alpha_i \in [0, 1]$ of the extra project revenue upon success, and an up-front transfer F_i. For budget balance reasons, we need $v_1 + v_2 = B$, $\alpha_1 + \alpha_2 = 1$, and $F_1 + F_2 = 0$. The restriction $\alpha_i \in [0, 1]$ is in fact a no-wager condition: absent this condition, the players would bet on the state and – in doing so – generate infinite utility. This no-wager condition can be derived endogenously by giving each player the ability to sabotage the project, i.e., by giving each player the ability to make sure that the project fails (Van den Steen 2005b). In that case, any contract with $\alpha_i \notin [0, 1]$ would give one of the players a strict incentive to sabotage the project. Anticipating that, the other would never accept the 'bet'. To maintain generality and to simplify the analysis, I simply impose the condition as an assumption. In stage 1c, the players' beliefs get privately realized.

In stage 2a, each player can send a cheap talk message from the set $\{X, Y\}$ to the other player. (As will become clear in section 3, in equilibrium only one party – the boss – will send a cheap-talk message and it will be interpreted by the other party – the subordinate – as a non-binding order.) In order to exclude non-interesting cheap talk equilibria, I will assume that the players select the Pareto-dominant equilibrium if one exists. Moreover, while sending messages is costless, I assume that players have a lexicographic preference for being obeyed: when they are payoff-indifferent, they will send a message only if it is strictly more likely to be obeyed than to be disobeyed. In step 2b, the players publicly choose their actions. Since the players' decisions are non-contractible, each agent always chooses the decision that is best from his perspective, given his beliefs and the contract negotiated in stage 1. (In this sense, an order from the boss does not directly constrain the subordinate's behavior.)

In step 2c, with probability p, each player can end the contract (before the game moves to stage 3). I will discuss the interpretation of the probability p immediately. (As will become clear in section 3, the ability to terminate the contract 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.) If either player ends the contract at this point, then the project is over and each player P_i gets his outside option $\underline{u}_i$. Player P_i's outside option will depend on the set of assets he owns (denoted A_i). Let P_i's outside option then be denoted $\underline{u}_i(A_i)$ and assume that both players' outside options depend in the same way on the set of assets they own, i.e., $\underline{u}_i(A) = \underline{u}(A)$. In this section, I take the outside options as exogenous. I will assume, first of all, that $\underline{u}(\emptyset) = 0$, $\underline{u}(\{a_1\}) = \underline{u}(\{a_2\}) > 0$, and $\underline{u}(\{a_1\}) + \underline{u}(\{a_2\}) = \underline{u}(\{a_1, a_2\}) < B + \nu$. All these conditions will arise naturally when the outside options are endogenized – as in section 5 – as future opportunities to match and execute the project with other players. I will furthermore assume that $\underline{u}(\{a_1, a_2\}) > B + q$, which I will derive in section 5 from a more primitive assumption (on the discount factor). I will discuss the meaning and role of this assumption at that time.

In period 3a, the state gets realized and the assets are committed (so that now $A_i = \emptyset$ for both players). In stage 3b, for sure, each player can end the contract. Note that the combined idea of stages 2c and 3b is that a player can always end the contract, but only with probability p is he or she able to do it in time before the assets are committed and the state revealed. I will assume that $p < 2 - \frac{1}{\nu}$. This assumption will have the effect to limit the probability that the principal can fire the agent prior to the assets being committed, which will then give the agent a temptation to disobey. If either player ends the contract in stage 3b, then the game is over and both players get their outside option, which is now 0 (since $A_i = \emptyset$).

In period 4a, both players can exert effort. As mentioned earlier, such effort affects the probability η that a success gives a payoff $B + 1$ rather than B. In particular, I will assume that $\eta = 1$ if both players exert effort and $\eta = \theta$ otherwise. Effort carries a private cost e for each player. I will again consider only Pareto-optimal outcomes.[11] In period 4b, finally, the payoffs are realized and the contract terms (v_i, α_i) get executed. Apart from their share in the residual income, both players also get a private benefit from success $\gamma \geq 0$, which captures intrinsic motivation.[12]

The reason to introduce – in stage 4 – moral hazard and intrinsic motivation is not to study their effect in full generality but rather to show that the interpersonal authority results of the model are robust to some amount of moral hazard and intrinsic motivation. I will therefore put explicit limits on their importance to keep the analysis simple. I distinguish two cases. The first case is one where moral hazard plays no role. In particular, with $\epsilon = \min(B + \nu - \underline{u}, \underline{u} - B - q, B) > 0$, the first case consists of the following assumption:

Assumption 1a $\theta = 1$ *and* $\gamma \leq \frac{p\epsilon}{4}$

The second case is when the moral hazard issue is non-trivial, which consists of the following assumption:

Assumption 1b $0 < (1 - \theta) \leq \frac{\epsilon}{2}$, $e \leq \frac{(1-\theta)^2}{2}$, *and* $\gamma \leq \frac{pe}{(1-\theta)}$.

where $\gamma \leq \frac{pe}{(1-\theta)}$ implies the earlier $\gamma \leq \frac{p\epsilon}{4}$. Unless explicitly mentioned, I will consider the second case, with the moral hazard issues.

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.

The purpose of the analysis is to determine the allocation of assets – and the equilibrium that goes with it – that will maximize the joint utility of all players, U. In doing so, I will consider only pure-strategy subgame-perfect equilibria that are not Pareto dominated. Such equilibria always

[11] In this case, any sequencing of the effort choices will automatically lead to the Pareto-optimal outcome.

[12] Note that with $\gamma > 0$, we need a further modification – beyond the ability to sabotage – to endogenize the condition $\alpha \in [0, 1]$. It is straightforward to see that a probabilistic structure for γ would do. Consider, for example, the case that $\gamma = 0$ with probability s and $\gamma = \hat{\gamma}$ with probability $1 - s$, with $\hat{\gamma}$ being 'sufficiently small'. Any $\alpha \notin [0, 1]$ would then lead to sabotage with probability (at least) s.

exist.[13] I will also consider whether the allocation would be different if the assets were traded or auctioned off at the start of the game.

3 Interpersonal Authority, Ownership, and Contracts

This section derives the main results of the paper. In particular, it shows that the optimal asset allocation is such that all assets required for the project are owned by one party; that the owner then hires others under a low-powered or fixed-wage contract to work on the project; and that these 'employees' take orders from the owner.

Before going to the formal analysis, let me give an overview of the different forces at work. As is clear from section 2, the unit of analysis is a project: a revenue stream that requires assets and depends on decisions. A basic result of the earlier literature is that – with differing priors – it is optimal to concentrate all income and control rights of a project in one hand (Van den Steen 2006a): as a person gets more control rights, she values (by revealed preference) income rights higher, so it is optimal to give her more income rights; as a person gets more income rights, she values control rights higher, making it optimal to give her more control rights.[14]

With income rights perfectly contractible, the issue is how to also move control rights around. As mentioned before, the approach will be to use an efficiency-wage type scheme: one player promises the other a high wage, tells him what to do, and threatens to fire him if he disobeys. To make this work, the principal must make sure not only that getting fired is costly to the agent, but also that he himself is committed to firing a disobeying agent. If not, he may be faced with the well-known situation of an obstructive employee who knows he can't get fired.

The rest of the analysis is about two mechanisms that strengthen interpersonal authority in such an efficiency-wage type context. The first mechanism – based on Van den Steen (2005b) – 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.

The second mechanism to strengthen the manager's interpersonal authority is through the allocation of asset ownership. Following the property rights theory, a key characteristic of ownership is residual control over the assets: when cooperation breaks down, ownership determines who gets the assets and thus what each player's outside option will be. On their turn, these outside options determine the cost of firing and getting fired, and thus the agent's incentives to obey the principal. There are actually two effects. First, if an employee owns an asset that is critical to the firm, then the firm will think twice before firing him and the employee will know that and use it. Second, the employee will care less about getting fired when he owns the asset since he has better outside options. The opposite is true when the firm owns all the relevant assets: the employee will feel both replaceable and vulnerable, and will thus be quick to please his boss.

I will now derive these results formally. Since interpersonal authority is such an important part of the analysis, I will first derive the conditions under which one participant obeys the other. To

[13] The focus on pure-strategy equilibria that are not Pareto dominated excludes only equilibria that are extremely similar to the ones I obtain here (where, for example, a player doesn't quit for sure, but with 'sufficiently high probability') and equilibria that are not very realistic or not very interesting (where, for example, both players quit simply because the other quits and therefore they are indifferent between quitting or not).

[14] As argued in Van den Steen (2006a), this *intuition* is specific to differing priors. First, with common priors, all players value residual income identically (in expectation). Second, an increase in a player's share of residual income does not, in itself, make him value control more in a context with common priors: it just makes him care more about making sure that the person with the most information gets to make the decision, whoever that is.

that end, I will study the subgame that starts in 2a – thus taking the asset allocation and the compensation contract (α_i, v_i) as given – and consider under what conditions one player will do what the other tells him to do. To simplify some of the notation, I will assume that the players are renamed such that P_1 physically gets the income of the project and pays P_2 a wage w and a share α of the extra revenue upon success, so $\alpha_1 = (1-\alpha)$ and $\alpha_2 = \alpha$. (Since this is just notation, it will not affect the results.) I will furthermore use Z_i to denote the action that player P_i believes is most likely to succeed; U to denote the joint expected utility of both players. Note that the expected project payoff according to player P_i when Z_i was implemented for both decisions is ν, while the expected project payoff according to P_i when Z_i was implemented for one decision but not for the other is q.

The following lemma describes the pure-strategy equilibria of the game (that are not Pareto dominated), starting in stage 2. I will use 'Authority by P_i' to denote the following equilibrium outcome: P_i orders P_j what to do; P_j obeys P_i's orders; P_i ends the contract (i.e., 'fires' P_j) if P_j were to disobey; P_i himself chooses Z_i; and neither player quits in equilibrium. I will use 'No Authority-Stay' to denote the following equilibrium outcome: neither player orders the other what to do, each player P_i chooses Z_i, and both players stay. If will, finally, use 'No Authority-Quit' to denote the following equilibrium outcome: neither player orders the other what to do, each player P_i chooses Z_i, and one or both players end the contract when they turn out to disagree. The following lemma then states all possible equilibria where the project gets executed and no player quits for sure. To state the lemma, let $\tilde{\alpha}_i = (\alpha_i + \gamma) - e$ if both players exert effort and $\tilde{\alpha}_i = (\alpha_i + \gamma)\theta$ if the players do not exert effort.

Lemma 1 *There exist only 4 pure-strategy (subgame) equilibria that are not Pareto dominated and where players do not quit for sure in equilibrium:*

1. *'Authority by P_1' if* $w \in [\max\left(B + \tilde{\alpha}_1 q - \underline{u}_1, \tilde{\alpha}_2(\frac{(1-p)}{p}(2\nu_2 - 1) - (1-\nu_2)) + \underline{u}_2, 0\right), \min(B + \tilde{\alpha}_1\nu_1 - \underline{u}_1, B)]$*, which gives* $U_{Au-1} = B + \tilde{\alpha}_1\nu_1 + \tilde{\alpha}_2\frac{1}{2}$,

2. *'Authority by P_2' if* $w \in [\max(-\tilde{\alpha}_2\nu_2 + \underline{u}_2, 0), \min\left(-\tilde{\alpha}_2 q + \underline{u}_2, B - \tilde{\alpha}_1\left(\kappa(2\nu_1 - 1) - (1-\nu_1)\right) - \underline{u}_1, B\right)]$*, which gives* $U_{Au-2} = B + \tilde{\alpha}_2\nu_2 + \tilde{\alpha}_1\frac{1}{2}$,

3. *'No Authority-Quit' if* $w \in [\max(-\tilde{\alpha}_2\nu_2 + \underline{u}_2, B + \tilde{\alpha}_1 q - \underline{u}_1), \min(B + \tilde{\alpha}_1\nu_1 - \underline{u}_1, \tilde{\alpha}_2(\kappa(2\nu_2 - 1) - (1-\nu_2)) + \underline{u}_2)] \cap [0, B]$ *or if* $w \in [\max(-\tilde{\alpha}_2\nu_2 + \underline{u}_2, B - \tilde{\alpha}_1\left(\kappa(2\nu_1 - 1) - (1-\nu_1)\right) - \underline{u}_1), \min\left(-\tilde{\alpha}_2 q + \underline{u}_2, B + \tilde{\alpha}_1\nu_1 - \underline{u}_1\right)] \cap [0, B]$*, which gives* $U_{NAu-Quit} = \frac{B + ((1+2\gamma)\eta - 2\epsilon)\nu_1}{2} + p\frac{\underline{u}_1 + \underline{u}_2}{2} + (1-p)\frac{B + ((1+2\gamma)\eta - 2\epsilon)q}{2}$.

4. *'No Authority-Stay' if* $w \in [-\tilde{\alpha}_2 q + \underline{u}_2, B + \tilde{\alpha}_1 q - \underline{u}_1] \cap [0, B]$*, which gives* $U_{NAu-Stay} = B + ((1+2\gamma)\eta - 2\epsilon)\frac{\nu_1 + q}{2}$.

Proof : See appendix. ■

The proof of the lemma is in appendix since doing the full backwards induction is uninstructive and quite tedious. Moreover, the basic intuition can be readily understood from looking at the different sets of conditions. Consider first the conditions for 'Authority by P_1': $w \geq B + \tilde{\alpha}_1 q - \underline{u}_1$ commits P_1 to firing P_2 when the latter disobeys[15], $w \geq \tilde{\alpha}_2(\frac{(1-p)}{p}(2\nu_2 - 1) - (1-\nu_2)) + \underline{u}_2$ makes it incentive compatible for P_2 to obey if P_1 fires him otherwise, $w \in [0, \min(B + \tilde{\alpha}_1\nu_1 - \underline{u}_1, B)]$ makes it ex-ante and ex-post individually rational to participate in the project. The conditions

[15] This is somewhat similar to the role of wages in Kahn and Huberman's (1988) model of up-or-out contracts.

for the 'No authority-Stay' case guarantee that neither player will quit if the other disobeys (as long as he himself makes the decision that he considers best), while the conditions for the the 'No authority-Quit' case guarantee that each player will do as he likes (even if the other quits upon disobedience) and that players quit if and only if they disagree.

I now turn to the main result of this paper, i.e., the result on ownership, fixed wages, and interpersonal authority. To state this result formally, let O_{ij} denote the ownership structure where asset a_1 is owned by P_i and asset a_2 is owned by P_j, with potentially $i = j$. Also remember that U denotes the joint expected utility, which is the objective function for the asset allocation.

The following proposition then says that allocating both assets to one player is the only ownership allocation that maximizes U for all parameter values. Moreover, the only equilibrium that then maximizes U (for all parameter values) is such that: residual income gets allocated as much as possible to the owner; the owner hires others under a contract that either pays a fixed wage or, sometimes, the minimal necessary incentives to get the employee to exert effort; the owner orders these other players what to do; and they obey. In other words, owners hire non-owners as employees, and these employees take their orders from the owner. I will discuss some more of the intuition after the proposition.

Proposition 1

- *An ownership allocation O maximizes U for all B iff $O \in \{O_{11}, O_{22}\}$.*
- *When the ownership structure is O_{ii}, the unique equilibrium is Au-P_i.*
- *Under assumption 1a, the contract in this equilibrium sets $\hat{\alpha} = 0$ and $w = \gamma(\kappa(2\nu_2 - 1) - (1 - \nu_2))$.*
- *Under assumption 1b, for any set of parameters excluding p, there exists some $\tilde{p} < 2 - \frac{1}{\nu_2}$ such that the contract in this equilibrium is as follows:*
  - *for $p \leq \tilde{p}$, the contract sets $\hat{\alpha} = 0$ and $w = \gamma(\kappa(2\nu_2 - 1) - (1 - \nu_2))$, and neither player exerts effort,*
  - *for $\tilde{p} < p \leq 2 - \frac{1}{\nu_2}$, the contract sets $\hat{\alpha} = \frac{e}{(1-\theta)} - \gamma$ and $w = \frac{e}{(1-\theta)}(\kappa(2\nu_2 - 1) - (1 - \nu_2))$, and both players exert effort.*

Proof : Let $\hat{U}_{\text{e}}$ and $\hat{U}_{\text{ne}}$ denote the maximal total utility of an authority equilibrium respectively with and without effort and $\hat{U}_{NAu}$ the maximal total utility of the respective no-authority equilibrium. Lemma 1 then showed that in the absence of w-feasibility constraints, $\hat{U}_{\text{ne}} > \max(\hat{U}_{NAu-Stay}, \hat{U}_{NAu-Quit})$ and (in the case with moral hazard, i.e., assumption 1b) $\hat{U}_{\text{e}} \geq \hat{U}_{\text{ne}}$. It follows that it suffices to show that $\hat{U}_{\text{ne}}$ is always feasible to conclude that any equilibrium must be *Au-P_i*.
To show this, consider $O = O_{11}$ (so that $\underline{u}_1 = \underline{u}$ and $\underline{u}_2 = 0$). An *Au-P_1* equilibrium then requires that

$$w \in [\max(0, B + \tilde{\alpha}_1 q - \underline{u}, \tilde{\alpha}_2(\kappa(2\nu_2 - 1) - (1 - \nu_2))), \min(B + \tilde{\alpha}_1 \nu_1 - \underline{u}, B)]$$

Since by assumption $B + \tilde{\alpha}_1 q < \underline{u}$ and $\kappa(2\nu_2 - 1) - (1 - \nu_2) > 0$, increasing α_1 increases the w-interval in the strong set-order (i.e, intervals at lower values are subsets of the intervals at higher values). Since increasing α_1 thus both increases the objective (U_{Au-1}) and loosens the constraint, any *Au-P_1* equilibrium must have either $\alpha_1 = 1$ (with neither player exerting effort) or $\alpha_1 = 1 - \left(\frac{e}{(1-\theta)} - \gamma\right)$ (with both players exerting effort). For further reference, the fact that $\hat{U}_{\text{e}} > \hat{U}_{\text{ne}}$ (in the case with moral hazard) implies that an *Au-P_i* equilibrium will have both players exerting effort whenever such effort equilibrium is feasible. To show now that $\hat{U}_{\text{ne}}$ is always feasible, note that the *Au-P_1* equilibrium without effort thus requires $w \in [\gamma(\kappa(2\nu_2 - 1) - (1 - \nu_2)), \min(B + (1 + \gamma)\nu_1 - \underline{u}, B)]$. Given the assumption that $\gamma \leq \frac{p\epsilon}{4}$, this condition can always be satisfied by $w = \gamma(\kappa(2\nu_2 - 1) - (1 - \nu_2))$, so that '*Au-$P_1$* without effort' with $\alpha_1 = 1$ is indeed always feasible. It also follows that the optimal equilibrium must always be an Au-P_i equilibrium (since any NAu equilibrium is dominated by a feasible equilibrium). This establishes the second point of the proposition and also reduces the overall problem to finding the optimal Au-P_i equilibrium.

I concluded above that the Au-P_i equilibrium will have both players exerting effort whenever the w-condition of such equilibrium can be satisfied. The w-condition for an equilibrium with effort requires

$$w \in [\left(\frac{e}{(1-\theta)} - e\right)(\frac{(1-p)}{p}(2\nu_2 - 1) - (1-\nu_2)), \min((1 + 2\gamma - \frac{2-\theta}{(1-\theta)}e)\nu_1 + B - \underline{u}, B)]$$

Since $\frac{(1-p)}{p}(2\nu_2 - 1) - (1-\nu_2) > 0$ (by the assumption that $p < 2 - \frac{1}{\nu_2}$) and since $\lim_{p\downarrow 0} \frac{(1-p)}{p} = \infty$, there exists some $\hat{p}$ such that the Au-P_1 equilibrium with effort is feasible only when $p \geq \hat{p}$. This implies the third part of the proposition.

All that is left to show is that only $O \in \{O_{11}, O_{22}\}$ maximize U for all B. To see this, consider $O = O_{ij}$ with $i \neq j$ and $B < \frac{u}{2}$. The feasibility condition for Au-P_1 (which was shown above to be the only equilibrium that maximizes U) is then

$$w \in [\max\left(0, B + \tilde{\alpha}_1 q - \frac{u}{2}, \tilde{\alpha}_2(\kappa(2\nu_2 - 1) - (1-\nu_2)) + \frac{u}{2}\right), \min(B + \tilde{\alpha}_1\nu_1 - \frac{u}{2}, B)]$$

which is impossible with $B < \frac{u}{2}$. This completes the proposition. ■

This proposition thus delivers the results on ownership, fixed wages, and interpersonal authority. (I will show, in the updated version of this paper, that there will always be effort in equilibrium when the decisions are contractible. This confirms the argument that it is indeed the need to exert authority that leads to lower-powered incentives and thus to less effort.) There are a few points of intuition that are important to mention. First of all, it is important to notice that allowing the players to end the contract after the realization of the state effectively limits the fixed wage that a player can promise to B. This wage limitation plays a key role. For example, if there is truly no limit on the wage you can promise, then you can always get obedience by promising an infinitely high wage and threatening to fire upon disobedience. In practice, with an arbitrarily high wage, the other party will try to get out of the contract when the project fails. That is exactly what the model captures. Another way to get the same result is to assume that players are capital constrained but can borrow from a bank with neutral beliefs, where the bank must make non-negative expected profits. Overall, this assumption thus reflects a real constraint.

Another important remark is that there is a second reason, beyond the need for authority, that makes it optimal in this model to allocate the assets to the person who also has control. The person who owns the assets must be compensated for their use (because otherwise he ends the project and takes the outside option). When the required compensation exceeds the feasible wage, then part of that compensation must be paid as a share of residual income. Apart from the fact that doing so will weaken the principal's authority, it also allocates residual income away from the person in control, which I argued to be inefficient: the person in control has a higher valuation of the residual income (by revealed preference) so it is efficient to allocate residual income as much as possible to him. This effect will persist even when decisions are contractible.

To address some common concerns, I now consider informally two variations on the main model: endogenous asset allocation and the case where moral hazard is really important.

Allocating assets by auction or trade The analysis simply determined the asset allocation that maximizes the joint utility of all players. A more elaborate model could endogenize this asset allocation process as a non-cooperative game. It turns out that – also for the model in section 5 with many assets and players – most traditional allocation processes, such as the 'efficient' multi-person bargaining process of Gul (1989) or an ascending-price auction, would result in exactly this utility-maximizing outcome.

This is, however, not a trivial result. It does not hold, for example, in Gans (2005) or in the private benefits model of Van den Steen (2006a). The difference between these settings and the current model is that, in the first of these settings, private valuations do not align with efficient allocations, and, in the second, non-owners care about the beliefs or preferences of owners. The robustness of this result in the current context is thus a topic for further research.

Moral hazard The model limited the importance of moral hazard to keep the analysis simple and focused on firms. Van den Steen (2005b) – which considers this issue in a closely related setting but without putting it in a theory-of-the-firm context and without assets – gives some indications of what would happen when moral hazard becomes much more important. In particular, it suggests that at some point there is a structural change in the equilibrium outcome from an 'authority' to a 'no authority' equilibrium. In the 'no authority' equilibrium, there is no more efficiency wage and no more orders or obedience, while the residual income is more shared as in the case of a partnership or independent agent. This issue requires more study in the current context.

4 Discussion

4.1 The Firm as a Legal Person

In the model and formal analysis, I followed Knight (1921) or Coase (1937) by not distinguishing formally between the firm and its entrepreneur-owner. Nevertheless, this model is (implicitly) built on a very clear definition of the firm that draws a definite line between the firm and its owner(s). In particular, I define the firm as a legal entity or legal person.[16]

Being a legal person means that a firm is a legal fiction that has all the standing and abilities of a physical person, i.e., the firm is an 'as if' person: it can own assets; it can write contracts; it has rights and obligations; etc. There are two important differences between a firm and a physical person. First, a firm is owned by shareholders. Nevertheless, it is distinct from these shareholders and indivisible. No shareholder directly owns any assets of the firm or can unilaterally decide to take out assets. The second key difference is that a firm cannot act on its own. Instead, its powers are exercised by a manager appointed by the shareholders. Even though the manager signs contracts for the firm, these contracts bind the firm rather than the manager.

In this 'personal' theory of the firm, integration simply means one legal person instead of two. Section 7 touches upon – and Van den Steen (2006b) discusses in detail – this issue, what it implies for the role of firm boundaries, and further implications.

The firm as a legal entity is straightforward to model explicitly. One can simply introduce a new type of player in the game – the firm – which is owned by other players who appoint a physical player to 'manage' the firm and exercise its rights. When, however, the model is limited – as in this paper – to a single enterpreneur-owner who does not sell his firm, the legal person of the firm is indistinguishable from the physical person of the owner. For transparency reasons, the current paper therefore does not make the distinction explicit. Van den Steen (2006b), however, uses this definition to study firms with a broader shareholder base, where such distinction is really necessary.

[16] This is the legal definition of a firm. My point (in particular in Van den Steen (2006b)) is that this legal definition is actually also a good starting point for an economic analysis of the firm.

One central focus of that paper is to show that ownership by a legal entity can strictly dominate direct individual ownership.

4.2 Knight's Theory of Entrepreneurship

It is difficult to avoid a comparison of this theory of the firm with Knight's (1926) theory of entrepreneurship since the two have important themes in common.

The starting point is the observation that differing priors is one way to interpret Knight's 'uncertainty' (as opposed to 'risk').[17] According to Knight, "(t)he conception of an objectively measurable probability [...] is simply inapplicable [to 'uncertainty']. The confusion arises from the fact that we do estimate the value or validity or dependability of our opinions or estimates, and such an estimate has the same *form* as a probability judgement [...]"(p.231). He also suggests to designate risk and uncertainty by "the terms 'objective' and 'subjective' probability"(p.233). These statements are consistent with an interpretation of 'uncertainty' and 'estimate' in terms of prior beliefs: your prior is your personal subjective estimate of some (currently) unknowable probability.

According to Knight, then, people differ (with respect to uncertainty) in the confidence in their judgment, which he described above as the 'estimate [of the] dependability of [one's] estimates' and which is captured here by a player's confidence ν. Knight then describes "the most fundamental change of all in the form of organization" as the "system under which the confident (...) 'insure' the doubtful and timid by guaranteeing to the latter a specified income in return for an assignment of the actual results"(p.269). While this statement is often interpreted as referring to risk-neutral people insuring the risk-averse, such interpretation actually does not seem consistent with Knight's interpretation of confidence, which refers to the probability that one's estimate is correct and is thus clearly something different than risk neutrality. Moveover, the terms 'doubtful and timid' also don't describe risk-aversion very well. Once we exclude risk-aversion, then one interpretation of Knight's theory of entrepreneurship is exactly the form of employment described in this paper: the principal tells the agent what to do and – in exchange – gives the agent a fixed wage, i.e., he *insures* the agent against the principal's mistakes. Note that in an extension of this paper where people differ in their confidence, it are indeed the more confident players who become the boss.

4.3 The Power of Interpersonal Authority

Economists often consider prices and contracting, rather than authority-like arrangements, as the default method to get things done. As Weitzman (1974) pointed out, however, lay people are more likely to consider more centralized arrangements first. The analysis in this paper suggests one reason why so many cooperative relationships in society are governed by interpersonal authority: all you need is a sufficiently high wage and a non-trivial possibility to end the cooperation. This is often much simpler/cheaper than a contract with state-contingent actions via message games etc.[18] Of course, studying such contracts is useful from a theoretical perspective, to understand what can be achieved and how.

Note that this paper does *not* imply that interpersonal authority is limited to within-firm relationships. I only argued that centralized asset ownership and low-powered incentives strengthen

[17] While Knight himself does not provide any clear formal definition of the concepts, but describes them in extensive prose, people have tried to concisely capture his definition of risk as 'randomness with knowable probabilities' and of uncertainty as 'randomness with unknowable probabilities.'

[18] Other papers have taken this simplicity as given and studied further implications. Simon (1951), for example, studies when such interpersonal authority dominates direct contracting on actions. Wernerfelt (1997) builds on this simplicity of authority to develop a theory of firms versus markets.

interpersonal authority, not that they are necessary conditions. It is perfectly possible for a firm to have some degree of authority over other firms or over non-employees. This may also require leaving the obeying party some rents combined with a threat of termination, as in Klein and Leffler (1981), but this issue requires further study.

5 Endogenous Outside Options

The model in the previous sections – with exogenous outside options – is in fact a reduced form for a richer model with endogenous outside options that I will present now. Besides endogenizing the outside options, I will also allow more players and more assets.

(Apart from the fact that it takes care of some formal details for working with multiple projects, this section is not necessary to follow the analysis on firm boundaries in section 6.)

5.1 The Full Model

In order to endogenize the outside options, I will embed the model of section 2 in a larger multi-period game where players can rematch if they end the project prematurely (i.e., when the bargaining breaks down or when someone ends the contract). Apart from these outside options, only stage 1 will be affected in a substantial way. (The other stages change slightly in response to the larger number of players and assets.)

Formally, consider now an economy with I players and two sets of A assets each, where I assume $I \to \infty$ and $A < \infty$. Denote the two sets of assets as $\mathcal{A}_1$ and $\mathcal{A}_2$. A 'project' is still a revenue stream R_n– with n indexing the projects – that requires two players (denoted $P_{1,n}$ and $P_{2,n}$) and two assets (denoted $a_{1,n}$ and $a_{2,n}$) with now one asset from each type, i.e., $a_{k,n} \in \mathcal{A}_k$. Any two players with two assets can execute a project. As in section 2, the two participants make decisions $D_{1,n}$ and $D_{2,n}$, and the probability of success Q_n of project R_n depends on these decisions being correct. (The probability $Q_n \equiv 0$ if there are less than 2 players or less than two (different) assets. More players or more assets, on the other hand, do not increase the probability.) To simplify the analysis, I will consider only the case of assumption 1a with $\theta = 1$, i.e., without moral hazard.[19] In all other respects, the structure of the payoffs and beliefs are the same as in section 2. Let me thus turn to the timing, represented in figure 3.

As mentioned before, the first stage – with the contracting – is the part of the game that is affected most by this modification. In particular, with more assets and more players, the first stage is not only about negotiating a contract, but also about matching players and assets. To specify this, I will define a 'bargaining solution' to consist of

- A set of N projects, $\mathcal{R} = \{R_1, \ldots, R_n, \ldots, R_N\}$, with $N \leq A$.
- For each project R_n, a set of players $I_n \subset I$ who are involved in the project.
- For each involved player $i \in I_n$ a contract $(\alpha_{i,n}, w_{i,n}, F_{i,n})$ where $F_{i,n}$ is an up-front transfer, $\alpha_{i,n} \in [0, 1]$, and where the budget is balanced: $\sum_{i \in I_n} \alpha_{i,n} = 1$, $\sum_{i \in I_n} w_{i,n} = B$, $\sum_{i \in I_n} F_{i,n} = 0$.
- For each project, the two assets $(a_{1,n}, a_{2,n})$ that will be used. No asset can be used in more than one project. The owner of each assets must be in I_n.

[19]The equivalent analysis for the case with moral hazard will be available from the author. The analysis requires an extra modification of the game to ensure the existence of a stationary equilibrium.

- For each project, the two players $(P_{1,n}, P_{2,n})$ who take the two actions/decisions $(D_{1,n}, D_{2,n})$. No player can take actions/decisions in more than one project. Both players must be in I_n.
- For definiteness, any $j \in I_n$ must be either one of the $P_{i,n}$ or own one of the $a_{k,n}$ or have $(\alpha_{j,n}, w_{j,n}, F_{j,n}) \neq (0,0,0)$.

Instead of formulating an explicit bargaining protocol, I will use the (efficient) Shapley solution.[20] The outside options are defined below. As to the timing itself, the bargaining solution gets determined in stage 1a, immediately after which the payments $F_{i,n}$ get made.

Stage 2 and 3 are identical to the one in section 2, except for a few minor changes that relate to the fact that there are now multiple projects and more than 2 players. In particular, in step 2b, any player will now be able to send a cheap talk message from $\{X, Y\}$ to any player with whom he is involved in a project. In step 2c and 3b, each player can end the contract for any *particular* project. In other words, a player can end one contract in which he is involved, but continue in his other projects. If any involved player ends the contract in stage 3b, then the project fails. For any project in which a player ends the contract, the game moves immediately to stage 4b. The two decision makers of a project (and the decision makers only) get intrinsic motivation $\gamma \geq 0$ from the project's success.[21]

Stage 4b captures the outside options. In particular, the game returns to 1a but now with payoffs discounted by $\delta \in (0,1)$, effectively starting a new period. If the game ended prematurely, then the assets were preserved and the owner of the assets can use his assets in the next period. Since project execution commits the assets that were used in that particular project, the set of assets will decrease over time and the game effectively ends when all assets are committed.

I will now assume that $\delta > \frac{B+q}{B+\nu}$. This assumption says essentially that a player (who is a claimant) gets a better payoff from getting both decisions right next period than from getting only one decision right this period. This is the assumption that gives $\underline{u}(\{a_1, a_2\}) \geq q$ in section 2. Without this, it is sometimes impossible for a manager to commit to firing a disobeying employee. The essence of the results would hold without the assumption, but it would be necessary to distinguish different cases, and the complexity of the analysis would increase substantially. I will also assume the equivalent of assumption 1a, which can be written $\gamma \leq \frac{p\tilde{\epsilon}}{4}$ for $\tilde{\epsilon} = \min((1-\delta)(B+\nu), B)$. As before, I will focus on Pareto-efficient, pure-strategy equilibria in order to get meaningful cheap-talk and simplify the analysis.

5.2 Analysis

The result and the intuition of lemma 1 extend fairly directly to this setting. In particular, lemma 1 holds now on the level of a project, subject to two changes. First, all contract variables are made project specific: $\alpha_{i,n}$ and $w_{i,n}$ replace α_i and w_i, etc. Second, the outside options are also project-specific: they are the expected revenues of the players if only this particular project ends prematurely.

There is also an important extension to lemma 1 for the case that someone other than the two decision makers has authority. To that purpose, let 'Authority by i' (for $i \notin \{P_{1,n}, P_{2,n}\}$) denote the following equilibrium outcome: i orders both $P_{1,n}$ and $P_{2,n}$ what to do; both $P_{1,n}$ and $P_{2,n}$ obey i's orders; i ends the contract if either $P_{1,n}$ or $P_{2,n}$ were to disobey (thereby ending the project);

[20] There are some well-known bargaining protocols that implement the Shapley solution, such as Gul (1989).

[21] The implicit idea is that γ captures career benefits of decision makers. Alternative specifications are obviously possible and lead typically to small modifications of the results.

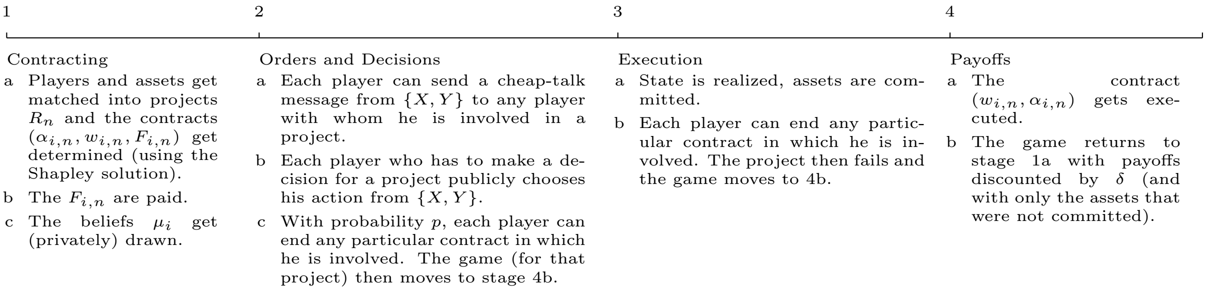


Figure 3: Time line of the model with endogenous outside options

and no participant quits in equilibrium. The following lemma then determines, by extension of lemma 1, the conditions for 'Authority by i'. Let $\check{\alpha}_j = \alpha_j + \gamma$ and remember that only decision makers $P_{1,n}$ and $P_{2,n}$ have intrinsic motivation γ.

Lemma 2 *Consider a project R_n with players j and k as decision makers. 'Authority by i' (with $i \neq j, k$) is an equilibrium if and only if all of the following conditions are satisfied:*

- $w_{j,n} + w_{k,n} \in [B + \alpha_{i,n} q - \underline{u}_{i,n}, \min(B + \alpha_{i,n}\nu_{i,n} - \underline{u}_{i,n}, B)]$
- $w_{j,n} \geq \check{\alpha}_{j,n} \left(\kappa(2\nu_{j,n} - 1) - (1 - \nu_{j,n})\right) + \underline{u}_{j,n}$ *and* $w_{k,n} \geq \check{\alpha}_{k,n} \left(\kappa(2\nu_{k,n} - 1) - (1 - \nu_{k,n})\right) + \underline{u}_{k,n}$

where $\kappa = (1 - p)/p$. Moreover, it is the unique (Pareto-efficient and pure-strategy) equilibrium on the interior of the interval. The joint utility $U_n = B + \alpha_{i,n}\nu + (1 + 2\gamma - \alpha_{i,n})\frac{1}{2}$.

Proof : From the proof of lemma 1, 'Authority by i' is an equilibrium if and only if the following conditions hold.

First, all players stay if j and k obey i. For i, this requires that $B + \alpha_{i,n}\nu_{i,n} - w_{j,n} - w_{k,n} \geq \underline{u}_{i,n}$ or $w_{j,n} + w_{k,n} \leq B + \alpha_{i,n}\nu_{i,n} - \underline{u}_{i,n}$. For j, this requires that $\check{\alpha}_{j,n}(1 - \nu_{j,n}) + w_{j,n} \geq \underline{u}_j$ or $w_{j,n} \geq -\check{\alpha}_{j,n}(1 - \nu_{j,n}) + \underline{u}_{j,n}$ and analogously for k: $w_{k,n} \geq -\check{\alpha}_{k,n}(1 - \nu_{k,n}) + \underline{u}_k$.

Second, all players stay even when the project turns out to be a failure. For i, this requires that $B - w_{j,n} - w_{k,n} \geq 0$ or $w_{j,n} + w_{k,n} \leq B$. The condition for j is that $w_{j,n} \geq 0$. The same holds for k.

Third, i quits if either j or k disobeys, which requires that $B + \alpha_{i,n} q - w_{j,n} - w_{k,n} \leq \underline{u}_{i,n}$ or $B + \alpha_{i,n} q - \underline{u}_{i,n} \leq w_{j,n} + w_{k,n}$.

Finally, both j and k obey (given that the other obeys). For j, this requires that

$$\check{\alpha}_{j,n}(1 - \nu_{j,n}) + w_{j,n} \geq p\underline{u}_{j,n} + (1 - p)(\check{\alpha}_{j,n}\nu_{j,n} + w_{j,n})$$

or

$$w_{j,n} \geq \check{\alpha}_{j,n} \left(\kappa(2\nu_{j,n} - 1) - (1 - \nu_{j,n})\right) + \underline{u}_{j,n}$$

and analogously for k. Note that this implies $w_{j,n} \geq -\check{\alpha}_{j,n}(1 - \nu_{j,n}) + \underline{u}_{j,n}$. Moreover, the assumption on p implies that $\kappa(2\nu_{j,n} - 1) - (1 - \nu_{j,n}) \geq 0$ and thus $w_{j,n} \geq 0$.

Putting things together, 'Authority by i' is an equilibrium if and only if the following conditions are satisfied:

- $w_{j,n} + w_{k,n} \in [B + \alpha_{i,n} q - \underline{u}_{i,n}, \min(B + \alpha_{i,n}\nu_{i,n} - \underline{u}_{i,n}, B)]$

- $w_{j,n} \geq \check{\alpha}_{j,n} \left(\kappa(2\nu_{j,n} - 1) - (1 - \nu_{j,n})\right) + \underline{u}_{j,n}$ and $w_{k,n} \geq \check{\alpha}_{k,n} \left(\kappa(2\nu_{k,n} - 1) - (1 - \nu_{k,n})\right) + \underline{u}_{k,n}$

Note also that – following the proof of lemma 1 – if these conditions are satisfied, then the unique (Pareto-efficient, pure-strategy) equilibrium is 'Authority by i' (except potentially in the end-points). This proves the proposition. ■

I now show that an appropriate version of proposition 1 also holds for this setting. To state that result formally, let o_i denote the set of assets owned by player i; $O = \{o_i\}_{i=1}^{\infty}$ a complete ownership structure; $\mathcal{O} = \{O : \cup_{i=1}^{\infty} o_i = \mathcal{A}_1 \cup \mathcal{A}_2; \forall i \neq j, o_i \cap o_j = \emptyset\}$ the set of feasible asset allocations; and $\overline{\mathcal{O}} = \{O \in \mathcal{O} : \forall i, \#(o_i \cap \mathcal{A}_1) = \#(o_i \cap \mathcal{A}_2)\}$ the ownership allocations such that the assets are owned in matching pairs, i.e., each player either owns no assets or owns exactly the assets that are necessary and sufficient for a set of projects. Moreover, let $n_i = \# o_i \cap \mathcal{A}_1$. Also remember that U denotes the joint expected utility, which is the objective function for the asset allocation.

The following proposition then says that allocating assets in matching pairs to players is the only ownership allocation that maximizes U for all parameter values. Moreover, the only equilibrium that then maximizes U for all parameter values is such that for each project both assets are owned by one player; all residual income gets allocated to that owner; owners hire others under a fixed-wage contract; the owner tells these other players what to do; and these non-owners obey. In other words, owners hire non-owners as employees and these employees take their orders from the owner. To clarify the proposition, I state the results in words between brackets.

Proposition 2 *An ownership allocation O maximizes U for all values of B iff $O \in \overline{\mathcal{O}}$. For any $O \in \overline{\mathcal{O}}$, the only equilibrium that maximizes U for all values of B is as follows:*

- *For any i with $n_i \geq 1$, $\exists \mathcal{R}_i \subset \mathcal{R}$ with $\#\mathcal{R}_i = n_i$ and s.t. $\forall R_n \in \mathcal{R}_i$: $i \in I_n$ and $a_{k,n} \in o_i$, and $\forall j \in \bigcup_{\mathcal{R}_i} I_n \setminus \{i\}$: $o_j = \emptyset$. (*For any player who owns assets, there exists a subset of the projects such that this player owns exactly all the assets for these projects.*)*
- *For some $R_{\hat{n}} \in \mathcal{R}_i$: $i \in \{P_{1,\hat{n}}, P_{2,\hat{n}}\}$. For all $R_n \in \mathcal{R}_i \setminus \{R_{\hat{n}}\}$: $\#I_n = 3$. (*Such player will be a decision maker on one of his projects. All his other projects have 3 participants.*)*
- *For all $R_n \in \mathcal{R}_i$, $\forall j \in I_n \setminus \{i\}$: $\alpha_{i,n} = 1$, $\alpha_{j,n} = 0$, $w_{j,n} = \gamma\left(\kappa(2\nu - 1) - (1 - \nu)\right)$, $F_{j,n} = w_{j,n}$. (*Such player will get the full residual income from all his projects and pay an efficiency wage to his employees to neutralize their temptation to disobey that comes from their intrinsic motivation.*)*
- *For all $R_n \in \mathcal{R}_i$, the equilibrium for the subgame starting in period 2 is 'Authority by i'.*
- *When $\gamma > 0$, then $n_i \leq 1$. (*When players have non-trivial intrinsic motivation, then all projects will be owned by different players.*)*

Proof : The proof will follow the same pattern as the proof of proposition 1. I will first show that the allocation of ownership and equilibrium proposed in the proposition do indeed maximize U. I will then show that for some values of B, it is the only allocation and equilibrium to do so.

Let U_n denote the joint expected utility that all players in I_n derive from R_n. As before, considering all possible equilibria implies $U_n \leq \overline{U}_{Au} = B + (1+\gamma)\nu_1 + \frac{\gamma}{2}$. Note, second, that any asset can be used only once for productive purposes. Third, delaying the execution of a project reduces the value generated due to discounting and the fact that the matching possibilities (may) go down. It follows that the maximum feasible value for U is $U \leq A\overline{U}_{Au}$.

Note that the proposed equilibrium does indeed generate $U = A\overline{U}_{Au}$, so it is a matter of showing that this is indeed an equilibrium when $O \in \overline{\mathcal{O}}$. Consider now first the outside options. Since this game starts with an equal number of assets of each type, and assets get committed in pairs, there will always be an equal number of assets of each type. With $I \to \infty$, the assets are the resources in short supply. In the limit as $I \to \infty$, the Shapley value depends then only on asset ownership, allocates all value to asset owners, and does so in proportion to the number of assets owned. It follows that for each pair of assets, i.e. in each particular project, a participant's outside option equals $\delta \overline{U}_{Au}$ if he owns the matching assets and 0 if he does not. (If the equilibrium generates $U_n = \overline{U}_{Au}$ for R_n, then the increase in outside value from owning one extra asset is $\delta \frac{\overline{U}_{Au}}{2}$.)

Since the projects are completely unrelated, it suffices to focus now on a single project. Consider first the case that i is not a decision-maker, and let the two decision-makers be j and k. The conditions for 'Authority by i' are now, given that in the proposed equilibrium $\alpha_{j,n} = \alpha_{k,n} = \underline{u}_{j,n} = \underline{u}_{k,n} = 0$ and $w_{j,n} = w_{k,n} = \gamma\left(\kappa(2\nu - 1) - (1 - \nu)\right)$

- $w_{j,n} + w_{k,n} \in [B + (1+\gamma)q - \underline{u}_{i,n}, \min(B + (1+\gamma)\nu_{i,n} - \underline{u}_{i,n}, B)]$
- $w_{j,n} \geq \gamma\left(\kappa(2\nu - 1) - (1 - \nu)\right)$ and $w_{k,n} \geq \gamma\left(\kappa(2\nu - 1) - (1 - \nu)\right)$

The last two conditions are trivially satisfied, while the first requires

$$2\gamma\left(\kappa(2\nu - 1) - (1 - \nu)\right) \in [B + (1+\gamma)q - \underline{u}_{i,n}, \min(B + (1+\gamma)\nu_{i,n} - \underline{u}_{i,n}, B)]$$

Since the assumption that $\delta(B + \nu) > B + q$ implies that $\delta(B + (1+\gamma)\nu) > B + (1+\gamma)q$ and thus $B + (1+\gamma)q - \underline{u}_{i,n} < 0$, it suffices to show that

$$2\frac{\gamma}{p} \leq \min(B + (1+\gamma)\nu_{i,n} - \delta(B + (1+\gamma)\nu + \frac{\gamma}{2}), B)$$

where I used the observation that $\kappa(2\nu - 1) - (1 - \nu) \leq \frac{1}{p}$. It thus suffices that

$$2\frac{\gamma}{p} + \frac{\delta\gamma}{2} \leq \min((1-\delta)(B + (1+\gamma)\nu), B)$$

which is implied by the assumption that $\gamma \leq \frac{p\tilde{\epsilon}}{4}$ for $\tilde{\epsilon} = \min((1-\delta)(B+\nu), B)$.

Consider next the case that i is himself a decision-maker on the project. Let the other decision maker be j. Note that this implies that $\underline{u}_{j,n} = 0$ while $\underline{u}_{i,n} = \delta\overline{U}_{Au}$. The condition for 'Authority by i' is then, given that in the proposed equilibrium $\alpha_{j,n} = 0$, $w_{j,n} = \gamma\left(\kappa(2\nu - 1) - (1 - \nu)\right)$,

$$w_{j,n} \in [\max\left(B + (1+\gamma)q - \underline{u}_{i,n}, \gamma(\kappa(2\nu_{j,n} - 1) - (1 - \nu_{j,n}))\right), \min(B + (1+\gamma)\nu_{i,n} - \underline{u}_{i,n}, B)]$$

which is satisfied by the earlier argument (since this is a weaker constraint).

I'm now left to show that this is the only allocation and the only equilibrium that maximize U for all values of B. To see this, note first that the above implies that $U = A\overline{U}_{Au}$ is always achievable. Furthermore, since the Shapley solution chooses a Pareto-efficient point in the feasible set, it will – in any subgame perfect equilibrium of an allocation that maximizes U– select a contract that implements $U_n = \overline{U}_{Au}$ for all projects. This implies, first of all, that (for $I \to \infty$) the outside options are $\delta\frac{\overline{U}_{Au}}{2}$ per asset that a player owns. It implies, second, that in any equilibrium, all projects must be executed in the first period and the subgame equilibrium for each project must be an 'Authority'-type equilibrium where the player with authority, say i, gets the full residual income, i.e. $\alpha_i = 1$. Note that once the bargaining solution is determined, each project is completely independent of the others. So I will focus on one project R_n and will argue that when $B \leq \frac{\delta(B+\nu)}{2}$, the solution proposed in the proposition is indeed the only one that implements 'Authority by i'. Any player – other than i– who owns an asset requires at least $w_{k,n} \geq \frac{\delta\overline{U}_{Au}}{2} \geq \frac{\delta(B+\nu)}{2}$ which is impossible to satisfy given the requirements that $\sum_{j \in I_n} w_{j,n} \leq B$ and $w_{j,n} \geq 0$. But that means that i must own the pair of matching assets that is used in R_n. Moreover, when $\gamma > 0$, then the situation where the asset owner is also a decision maker gives a higher utility ($U = B + (1+\gamma)\nu + \frac{\gamma}{2}$) than the situation where the asset owner is not a decision maker ($U = B + \nu + \gamma$). Aggregating these conditions over all projects imply the proposition. ■

The outside option of a player is now $\delta \overline{U}_{Au}/2$ times the number of assets that he or she owns. It follows that the conditions of section 2 that $\underline{u}(\emptyset) = 0$ and $0 < \underline{u}(\{a_1\}) + \underline{u}(\{a_2\}) = \underline{u}(\{a_1, a_2\}) < B + (1 + \gamma)\nu + \frac{\gamma}{2}$ arise indeed naturally from the model.[22] Furthermore, the assumption that $\delta(B + \nu) > B + q$ implies the earlier assumption that $\underline{u}(\{a_1, a_2\}) > q$. So the model literally endogenizes the outside options of the earlier model.

Note that in this outcome, a player can 'own' more than one project. If a player does own multiple projects, then employees handle all decisions on all these projects but one. It is still, however, the owner who tells each employee what to do. In the setting above, such multi-project ownership only happens when players have no intrinsic motivation and then generates exactly the same value as a solution where each project is owned by a different player. That will change in the next section, where such 'merger' really affects expected residual income.

6 Fundamental Disagreement, Break-up, and Firm Boundaries

One important purpose of a theory of the firm is to provide foundations for 'markets versus hierarchies' decisions. In this section, I use the theory of this paper to propose the risk of 'break-up' as a motivation for integration. In particular, I will show that fundamental disagreement may cause two (separate) firms to go their own way despite coordination being efficient from the perspective of any outsider. Integration can solve this issue by giving – in equilibrium – one manager interpersonal authority over all employees, and thus by eliminating disagreement among those in control.[23] In fact, I will show – in proposition 3c below – that integration can be strictly optimal *even when there is perfect ex-post Coasian bargaining over the decisions* (and even when there are no incontractible investments that cause holdup). This is in marked difference to the property rights theory of the firm (Grossman and Hart 1986, Hart and Moore 1990, Hart 1995) – where integration only matters in the presence of incontractible investments – and to Hart and Holmstrom (2002), Baker, Gibbons, and Murphy (2006), Hart and Moore (2006), and related work where ex-post incontractibility of the actions is key. I then further show that the risk of breakup gets leveraged into an important additional effect: the anticipation of such break-up may prevent relation-specific investments since the parties may fear that they will not reap the benefits of the investment. Proposition 3d shows that this may make integration strictly optimal *even when these relation-specific investments are ex-ante contractible* (and either with or without ex-post Coasian bargaining on the decisions). This is again in contrast to the literature which has focused on hold-up caused by non-contractible ex-ante investments. Situations with (potentially contractible) relation-specific investments are probably the most important application of this theory.

This theory of break-up is partially motivated by personal observations of real merger and acquisition decisions. In one case, for example, 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'. In particular, both firms would have to make considerable relation-specific investments (integrating the product lines) and there was a risk

[22] Note that the stated condition on the outside option was actually $\underline{u}(\{a_1, a_2\}) < B + \nu$. The condition on γ, however, made it sufficient to ensure that $\underline{u}(\{a_1, a_2\}) < B + (1 + \gamma)\nu + \frac{\gamma}{2}$.

[23] Analogous to the property rights theory, the only change will be a shift in asset ownership. All other 'changes', including the changes in contracts, are really changes in the equilibrium outcomes.

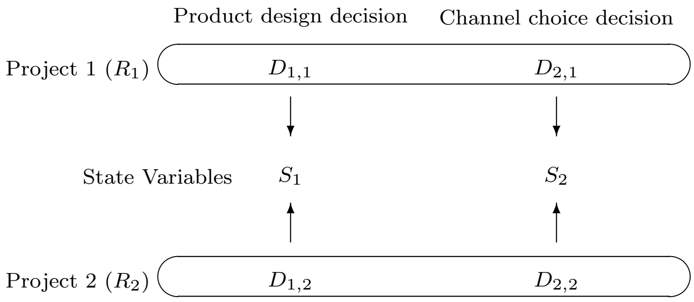


Figure 4: Projects, Actions, and States

that future disagreements could prove unresolvable and cause a break-up of the alliance. This issue was instrumental in the firm's decision to choose an acquisition.

While break-up has similarities to hold-up, it also has significant differences. In both cases, a key issue is that some players fear not to get the full return on their investment. While hold-up redistributes value, break-up causes some value not to be realized at all. One implication of this difference is that break-up is itself costly, while hold-up is costly only through its effect on investments. Another implication of this difference is that hold-up can be solved by ex-ante contractibility of the investments, but that that solution often does not work for break-up. In both senses, break-up is actually a harder issue.

Before going into the analysis, it is probably worth stressing what is unusual about this result, as captured most clearly in proposition 3c. Even when coordination between two players' actions is *always efficient* from an outsider's perspective and the players can *bargain perfectly* on the actions, they will often end up *not* coordinating. The reason is that due to their diverging beliefs, they cannot agree what to coordinate on and, from each firm's perspective, taking the wrong action is more costly that losing coordination (even if you can internalize the externality by contracting). While this may seem surprising in view of the Coase Theorem, it is due to a difference between objective and subjective efficiency (Van den Steen 2005a). It is this effect that causes firm integration to be optimal. The cost of integration in this model will be that one of the projects is not owned by its optimal owner when considered in isolation (i.e., it is not owned by the player with the strongest views about how to do this project). The benefit of integration is that it excludes the possibility of fundamental disagreement by giving all control to one player.

To analyze this issue, consider the following variation on the model of section 5. Assume that there are two possible projects with respective revenue streams R_1 and R_2, and that each project requires exactly one asset. Denote the asset required for project R_n as a_n, and assume that there is exactly one asset of each type. The question will be whether these two assets should be owned together or separately (i.e., whether the projects should be merged or not) and how that affects the contracts and control.

As before, each project R_n requires two participants, and each participant $P_{i,n}$ has to make a decision $D_{i,n} \in \{X_i, Y_i\}$. Also as before, one and only one of these decisions is correct, as captured by state variable $S_i \in \{X_i, Y_i\}$. Note that, as depicted in figure 4, the two decisions of one project n ($D_{1,n}$ and $D_{2,n}$) now have different state variables (S_1 and S_2). On the other hand, I will assume that the corresponding decisions of the two projects depend on the same state variable: so $D_{1,1}$ and $D_{1,2}$ have the same state variable S_1, and $D_{2,1}$ and $D_{2,2}$ have the same state variable S_2. A typical example of such situation is when R_1 and R_2 are new products in similar markets – say consumer electronics – and $D_{1,n}$ is a product design decision while $D_{2,n}$ is a channel decision. It will then often be the case that the success of both designs depend on the same set of factors (the evolution of taste, standards, etc.), while the success of both channel choices depends on another, though also common, set of factors (evolution of retailing, information technology, etc.). This connection between the two projects plays an important role in the analysis.

In order to make sure that it is sometimes strictly optimal for both projects to be owned by different people, I will assume that the two projects attach different weights to the two decisions, and will use parameter $\rho \in (0,1)$ to capture that. (I will present the functional form of the probabilities of success below, which will then clarify the formal role of ρ.) In the example above, the success of one product will be more sensitive to the design decision while the success of the second product will be more sensitive to the channel choice. Correspondingly, I will assume that some people have more confidence about the design ($D_{1,n}$) decisions, while others have more confidence about the channel ($D_{2,n}$) decisions. Formally, let μ_i^j denote player i's belief that $S_j = X_j$. In analogy to before, each μ_i^j will be an independent draw from a binary distribution that puts 50/50 probability on ν_i^j and $(1-\nu_i^j)$, for a given set of parameters $\nu_i^j \in (.5, 1)$. I will assume that the players are divided into two types. Players of type 1 have $\nu_i^1 = \overline{\nu} > \underline{\nu} = \nu_i^2$ while players of type 2 have $\nu_i^1 = \underline{\nu} < \overline{\nu} = \nu_i^2$. In other words, the type-1 players have stronger beliefs about the design ($D_{1,n}$) decisions while the type-2 players have stronger beliefs about the channel ($D_{2,n}$) decisions.

Finally – in order to make sure that it is sometimes strictly optimal for both projects to be owned by the same person – I will assume that the project success depends not only on the decisions being correct, but also on whether the decisions of both projects are coordinated. In particular, there will be some gain if, say, the channel choices of both projects are identical. One could, for example, imagine that the two products are complements (game consoles and games), so that using the same channel improves the chances of success. (On the design side, we could imagine that using matching styles or adopting the same connection standard would improve the probability of success of both projects.) I will parameterize the importance of such coordination by $\beta \in [0,1)$. It is this need for coordination that will create the conflict issues.

To express all these ideas formally, let $d_{i,n}$ be the indicator function that decision $D_{i,n}$ is correct, and let the respective probabilities of success (conditional on execution) for the two projects be $Q_1 = (1-\beta)\frac{d_{1,1}+\rho d_{2,1}}{2} + \beta I_{\{D_{2,1}=D_{2,2}\}}$ and $Q_2 = (1-\beta)\frac{\rho d_{1,2}+d_{2,2}}{2} + \beta I_{\{D_{2,1}=D_{2,2}\}}$. To keep the analysis tractable, I will consider the case without moral hazard, i.e., $\theta = 1$, and with negligeable intrinsic motivation, i.e., the limit where $\gamma \downarrow 0$. The earlier assumption that $\delta > \frac{B+q}{B+\nu}$ becomes a bit more complex due to the coordination issue and the different confidence levels: $(1-\delta) < \frac{\rho(\underline{\nu}-\frac{1}{2})-\frac{\beta}{(1-\beta)}}{(\frac{B}{(1-\beta)}+\frac{\overline{\nu}+\rho\underline{\nu}}{2})}$ (which obviously requires $\frac{\beta}{(1-\beta)} < \rho(\underline{\nu}-\frac{1}{2})$). Otherwise, the model is identical to that of section 5.

I now first establish that absent coordination issues ($\beta = 0$) the assets are optimally owned by players of different types. In other words – absent coordination issues – it is strictly optimal to have two separate firms. The reason why separate ownership is optimal is that, in terms of the earlier example, type-1 players are more confident about product design decisions and thus have a higher expected value from 'owning' the more design-dependent project, while type-2 players are more confident about channel design decisions and thus have a higher expected value from 'owning' the more channel-dependent project (where 'owning' a project is meant as having the project's full residual income and control, through asset ownership and contracting). Think Apple versus Dell. Since the projects are unrelated, separate ownership is thus optimal.

Proposition 3a *When $\beta = 0$, an allocation of asset ownership maximizes U for all values of B if and only if it allocates a_1 to a player of type 1 and a_2 to a player of type 2.*

Proof : Since – with $\beta = 0$ – there is no connection between the two projects, I can treat each project separately. The analysis is then essentially analogous to the proof of proposition 2.

One important difference is that for each project $U_n \leq B + \frac{\overline{\nu}+\rho\underline{\nu}}{2}$. This expected utility level (only) gets attained when for, say, project R_1 the equilibrium is 'Authority by i' with i a player of type 1.[24] A completely analogous argument to that of proposition 2 then implies that the (only) ownership structure that implements this equilibrium and thus reaches this expected utility for all values of B is one in which i owns asset a_1. Repeating this argument for R_2 implies that an ownership structure maximizes U (for all values of B) if and only if it allocates a_1 to a player of type 1 and a_2 to a player of type 2. ■

When there are coordination issues ($\beta > 0$), however, an integrated firms may be strictly optimal. In particular, the following proposition says that for intermediate values of β it is uniquely optimal to have only one owner. The intuition for this result is as follows. Consider the situation with two separate firms, and assume that the firms' managers disagree on the right course of action for decision $D_{2,n}$. Being different firms implies that each has full residual income and control for its project. As a consequence, a manager incurs a considerable cost from trying to coordinate since it requires him to make a decision that he deems suboptimal. Disagreement thus causes a conflict between the desire to make an optimal decision and the desire to coordinate. When one firm owns both assets, its manager always agrees with himself so that that conflict vanishes. That is exactly what integration accomplishes.

Proposition 3b *When* $\frac{(1-\rho)(\overline{\nu}-\underline{\nu})}{2} < \frac{\beta}{(1-\beta)} < \rho(\underline{\nu} - \frac{1}{2})$*, an allocation maximizes* U *for all values of* B *if and only if it allocates both assets to one player.*

Proof : An argument similar to before implies that there are two candidates for the optimal allocation, with accompanying equilibrium:

1. Asset a_n gets allocated to a type-n player. The owner (of project R_n), denoted $P_{1,n}$, makes a $(w, \alpha, F) = (0, 0, 0)$ offer to a non-owner, and the ensuing equilibrium is 'Authority by $P_{1,n}$'.
2. Both assets get allocated to one player. The owner, denoted P_1, makes $(w, \alpha, F) = (0, 0, 0)$ offers to three non-owners, and the ensuing equilibrium is 'Authority by P_1' (where P_1 quits the project if any of the three disobeys).

Consider now first the situation that both assets are allocated to one player. Since $D_{2,1}$ and $D_{2,2}$ have the same state variable, this player will always coordinate the decisions. It follows that the expected revenue is $2B + (1-\beta)(1+\rho)\frac{\overline{\nu}+\underline{\nu}}{2} + 2\beta$.

Consider next the case that a type-k player owns a_k. As long as $\beta < (1-\beta)\rho(\underline{\nu} - \frac{1}{2})$, neither owner is willing to choose an action he considers less likely to succeed in order to achieve coordination. It follows that the total expected payoff from split ownership equals $2B + 2(1-\beta)\frac{\overline{\nu}+\rho\underline{\nu}}{2} + \beta$, where the β comes from the fact that the players will agree, and thus automatically coordinate, half the time.

Finally, the merged firm gives higher total utility if $2B + (1-\beta)(1+\rho)\frac{\overline{\nu}+\underline{\nu}}{2} + 2\beta \geq 2B + 2(1-\beta)\frac{\overline{\nu}+\rho\underline{\nu}}{2} + \beta$ or $\frac{\beta}{(1-\beta)} \geq (1-\rho)\frac{\overline{\nu}-\underline{\nu}}{2}$ ■

Note that integration has two effects that together drive the result. First, in equilibrium, the integrated company will be the residual claimant on both projects. Since both projects are now evaluated by one and the same manager, there can be no disagreement on the optimal course of action. Second, in equilibrium, integration also concentrates the necessary control, i.e., the one

[24] To see that $B + \frac{\overline{\nu}+\rho\overline{\nu}}{2}$ can*not* be attained by giving authority over decision $D_{1,1}$ to a player of type 1 and authority over $D_{2,1}$ to a player of type 2, note that the project's expected returns according to the respective players are $B + \frac{\overline{\nu}+\rho\frac{1}{2}}{2}$ and $B + \frac{\frac{1}{2}+\rho\overline{\nu}}{2}$. No allocation of residual income can make the expected utility larger than $B + \frac{\overline{\nu}+\rho\underline{\nu}}{2}$.

manager has interpersonal authority over all employees in both projects and can thus implement the coordinated actions.

Ex-post bargaining and unresolvable disagreement The intuition also suggests that allowing ex-post Coasian bargaining over the decisions (i.e., bargaining at the time the decisions are made, after the uncertainty is realized) will not necessarily solve this issue: each firm remains the residual claimant on its project and thus bears the consequences of making a (in its eyes) suboptimal decision.[25] It follows that a firm will still be faced with a conflict between trying to make the right decision and trying to coordinate. In particular, the following proposition identifies conditions under which integration is still strictly optimal.

Proposition 3c *When the decisions $D_{i,n}$ are perfectly contractible at the end of step 2a but $\frac{(1-\rho)(\overline{\nu}-\underline{\nu})}{2} < \frac{\beta}{(1-\beta)} < \rho\frac{\underline{\nu}-\frac{1}{2}}{2}$, an allocation maximizes U (for all values of B) if and only if it allocates both assets to one player.*

Proof : Consider first the case that the assets are separately owned (with a k-type player owning a_k). If the owners would try to coordinate by contracting ex-post on the decision, they would contract such that the project that is least sensitive to the second (channel) decision, which is project R_1, follows the belief of the other owner. In other words, player $P_{2,1}$ implements what the owner of R_2 believes is best for $D_{2,1}$. This will be optimal if and only if $2\beta \geq (1-\beta)\rho(\underline{\nu} - q)$ where the LHS is the gain from coordination and the RHS is the cost of coordination to the owner who cares least about this particular decision (and who will thus be the one taking his less preferred action). So when $\frac{\beta}{(1-\beta)} < \rho\frac{\underline{\nu}-\frac{1}{2}}{2}$ then the separate firms will not coordinate when their managers disagree.

The condition for the integrated firm to generate higher U remains the same as in proposition 3b: $\frac{\beta}{(1-\beta)} \geq (1-\rho)\frac{\overline{\nu}-\underline{\nu}}{2}$. This concludes the proposition. ■

This is thus an 'unresolvable disagreement': even if decisions are contractible, the players will not come to an agreement and will lose the coordination benefits. The driving force here is that each firm remains the residual claimant on its own project and thus bears the full costs of following a suboptimal (from its manager's point of view) course of action. When this (subjective) cost is larger than the coordination benefits, then coordination will not happen. In the integrated firm, on the other hand, coordination always happens, as before. The issue here is thus not a failure of Coasian bargaining, as with private information, but a change in what is (subjectively) efficient.

Relation-specific Investments This break-up issue becomes even more significant when there are relation-specific investments. In particular, the anticipation of break-up may prevent such investments and thus strengthen the case for integration.

To see this formally, assume that getting the coordination benefit requires an up-front investment b by, say, the firm that owns asset a_1. (The coordination benefit, however, still only obtains if the firms actually coordinate). The following proposition then says that a merged firm is more likely to make the relation-specific investment, that integration can be strictly optimal, and that that still holds *even when* I allow perfect ex-ante contractibility of the investments and perfect ex-post Coasian bargaining on the decisions.

[25] The terminology 'ex-post' refers to the bargaining happening after the uncertainty has been realized (even though it is still the middle of the game). Alternative terminology seemed to make things more confusing or more complex than necessary.

Proposition 3d *The set of parameters for which the separate firms invest is a strict subset of the set of parameters for which the merged firm invests, whether or not the investment is ex-ante contractible and whether or not the decisions are contractible at the end of step 2a.*

Assume that $\frac{(1-\rho)(\overline{\nu}-\underline{\nu})}{2} < \frac{2\beta-b}{(1-\beta)}$, $\frac{\beta}{(1-\beta)} < \rho\frac{\underline{\nu}-\frac{1}{2}}{2}$, $\beta < b < 2\beta$, *the investment is perfectly contractible (prior to stage 1), and the decisions are contractible at the end of step 2a. In that case, an allocation maximizes U for all values of c if and only if it allocates both assets to one player.*

Proof : Consider first a merged firm. It will always coordinate and it will invest if $b \leq 2\beta$. Consider next two separate firms. From the earlier propositions, it follows that they will coordinate absent ex-post Coasian bargaining (on the decision) if $\frac{\beta}{(1-\beta)} \geq \rho(\underline{\nu} - \frac{1}{2})$ and with ex-post Coasian bargaining if $\frac{\beta}{(1-\beta)} \geq \rho\frac{\underline{\nu}-\frac{1}{2}}{2}$. Moreover, if they do not coordinate and the investment is not contractible, then the firm that owns a_1 invests if $b \leq \frac{\beta}{2}$. If they either coordinate and the investment is non-contractible or they do not coordinate but the investment is contractible, then they invest if $b \leq \beta$. Finally, if they do coordinate and the investment is contractible then they invest whenever the merged firm invests. It follows that whenever the separate firms invest, so will the merged firm. But there is a part of the parameter space (with non-empty interior) where the merged firm invests but the separate firms do not. This proves the first part of the proposition.

For the second part of the proposition, consider the case that the firms can contract on the investment and they can contract ex-post on the decision. By the earlier calculations, separate firms will not coordinate even if they can contract ex-post whenever $\frac{\beta}{(1-\beta)} < \rho\frac{\underline{\nu}-\frac{1}{2}}{2}$. Second, conditional on separate firms not coordinating upon disagreement, the above implied that a merged firm will invest but separate firms will not invest, even if the investment is contractible, whenever $\beta < b < 2\beta$. Finally, some algebra implies that the expected value from a merged firm that makes the investment is larger than the expected value of separate firms that do not make the investment when $\frac{(1-\rho)(\overline{\nu}-\underline{\nu})}{2} < \frac{2\beta-b}{(1-\beta)}$. This proves the proposition. ■

The reason why ex-ante contractibility of the investment cannot solve the problem is that the risk of break-up still remains. As long as the parties cannot commit to align their decisions in the future, the up-front investments are not worth making. Of course, the issue will be more important and apply more broadly when investments and ex-post decisions are not contractible.

Overall, the potential for future unresolvable disagreement and break-up can prevent relation-specific investments, even if the investments are perfectly contractible and even if there is perfect ex-post Coasian bargaining over the decisions. The prediction that the need for relation-specific investments leads to integration has found support in, for example, Monteverde and Teece (1982).

7 Conclusion

This paper presented a theory of the firm in which unified 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 its employees. This theory provides micro-foundations for the idea that bringing a project in a firm gives the manager interpersonal authority over the employees working on the project, which is a key implicit assumption in many theories of firm boundaries. I also argued that Knight's view of entrepreneurship can be interpreted along these lines.

This theory of the firm then allowed me to suggest and derive a new economic argument for integration between firms: the risk of break-up (due to fundamental disagreement) between independent firms – and the fact that the anticipation of such break-up may prevent relation-specific investments – may make integration optimal. This theory makes economic sense of the notion – often expressed by managers – that managers may prefer outright mergers and acquisitions because they feel they need full control over the other firm if they will heavily depend on it. I show that such integration makes sense *even when the actions are perfectly ex-post contractible.*

The underlying definition in this paper of the firm as a legal person is further developed in Van den Steen (2006b). That paper shows how this definition of the firm as a legal entity allows extending the current theory to a setting with multiple shareholders. It shows, in particular, how and why the firm's owning the assets can dominate each individual shareholder's directly owning part of the assets. In other words, it shows how 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. The paper also treats some related issues. One such issue is the role of firm boundaries. In particular, once a firm is defined as a legal person or legal entity, firm integration is defined as one legal person versus two. The integrated entity has all rights and obligations, including all contracts and ownership, of these two entities. The manager of the integrated entity is thus the 'as if' party to all the contracts and the 'as if' owner of all the assets. While it is possible – and useful – to enumerate a firm's rights and obligations (including ownership), translating these to strict and absolute boundaries is not necessary in this framework, given that it is already well-defined what it means to be one versus two firms. Moreover, a contract is never unambiguously 'part' of either party, since no party has residual control over the contract. In this framework, what matters is what the firm can do, rather than where it 'ends'. Other issues that the paper discusses are firms that own no assets and the role of contracts.

The current paper also suggested some other further research questions. Section 4 raised, apart from the definition of a firm as a legal entity, also the issue 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. Section 6, finally, raises the obvious issue of firm boundaries. Klepper and Thompson (2006) provide interesting evidence on the role of strategic disagreement in the formation of new firms through spinoffs.[26] I believe that the current theory raises some interesting conjectures in that direction.

[26] Although not stressed in the paper, the model assumes differing priors with regard to the precision of the information.

A Proofs

Proof of Lemma 1: For later purposes, I will distinguish in this proof the confidence of both players, ν_1 and ν_2. Furthermore, I will – if necessary – rename the players so that P_2 gets w and α (so P_1 gets $B - w$ and $(1 - \alpha)$).

Consider first stage 4a. A player P_i will exert effort if $(\alpha_i + \gamma)(1 - \theta) \geq e$ and he believes that the other player will also exert effort. Since we consider only equilibria that are not Pareto-dominated, both players will exert effort iff $(\alpha_i + \gamma)(1 - \theta) \geq e$ for both. Let now $\eta = 1$ and $\epsilon = e$ if $(\alpha_i + \gamma)(1 - \theta) \geq e$ for both players and $\eta = \theta$ and $\epsilon = 0$ otherwise. Let finally $\tilde{\alpha}_i = (\alpha_i + \gamma)\eta - \epsilon$, so that $\tilde{\alpha}_i$ is P_i's expected payoff from a success net of his/her potential cost of effort.

Consider next stage 3b. P_1 quits upon failure if $B - w < 0$ or $w > B$, and upon success if $B + \tilde{\alpha}_1 - w < 0$ or $w > B + \tilde{\alpha}_1$, while P_2 quits upon failure if $w < 0$ and upon success if $w + \tilde{\alpha}_2 < 0$ or $w < -\tilde{\alpha}_2$. Note that if $w \notin [0, B]$, then w gets only paid upon success. In what follows, I do the analysis assuming that $w \in [0, B]$ and then consider at the end the case that $w \notin [0, B]$.

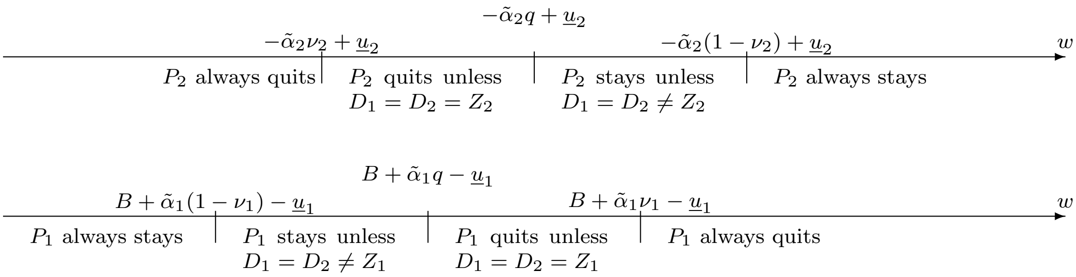
Consider now stage 2c. Let v_i denote P_i's fixed payment (independent of success or failure). If $D_i \neq D_j$ then P_i expects the project to have a payoff of $q\eta$. If i believes that $D_i = D_j = Z_i$, then $u_i = \tilde{\alpha}_i \nu_i + v_i$ and P_i stays if $\tilde{\alpha}_i \nu_i + v_i \geq \underline{u}_i$. If $D_i \neq D_j$ then $u_i = \tilde{\alpha}_i q + v_i$ and P_i stays if $\tilde{\alpha}_i q + v_i \geq \underline{u}_i$. If $D_i = D_j \neq Z_i$, then $u_i = \tilde{\alpha}_i(1 - \nu_i) + v_i$ and P_i stays if $\tilde{\alpha}_i(1 - \nu_i) + v_i \geq \underline{u}_i$.

Consider now, for example, player P_1. If $D_1 = D_2 = Z_1$, then he stays if $B + \tilde{\alpha}_1 \nu_1 - \underline{u}_1 \geq w$. If $D_1 \neq D_2$ then he stays if $B + \tilde{\alpha}_1 q - \underline{u}_1 \geq w$. If $D_1 = D_2 \neq Z_1$, then he stays if $B + \tilde{\alpha}_1(1 - \nu_1) - \underline{u}_1 \geq w$. In other words,

- if $w \leq B + \tilde{\alpha}_1(1 - \nu_1) - \underline{u}_1$ then he always stays
- if $B + \tilde{\alpha}_1(1 - \nu_1) - \underline{u}_1 \leq w \leq B + \tilde{\alpha}_1 q - \underline{u}_1$ then he stays unless the worst happens ($D_1 = D_2 \neq Z_1$)
- if $B + \tilde{\alpha}_1 q - \underline{u}_1 \leq w \leq B + \tilde{\alpha}_1 \nu_1 - \underline{u}_1$ then he stays only if the best happens ($D_1 = D_2 = Z_1$)
- if $B + \tilde{\alpha}_1 \nu_1 - \underline{u}_1 \leq w$ then he never stays

The conditions for P_2 are analogous.

This can be summarized graphically as follows, where the upper and lower graph may line up differently depending on $\underline{u}_1$ and $\underline{u}_2$.



In an abuse of notation, I will use – in what follows – $[a, b]$ to mean $[a, b]$ when $b \geq a$ and $\emptyset$ otherwise.

If $w \notin Z = [-\tilde{\alpha}_2 \nu_2 + \underline{u}_2, B + \tilde{\alpha}_1 \nu_1 - \underline{u}_1]$, then at least one player quits for sure and the payoff is simply $\underline{u}_1 + \underline{u}_2$. So let's consider now $w \in Z$.

Let now $X = [-\tilde{\alpha}_2 q + \underline{u}_2, B + \tilde{\alpha}_1 q - \underline{u}_1]$. Note that $X \subset Z$. If $w \in X \subset Z$, then neither player quits when the other disobeys (as long as he himself did what he thinks was best), so it is optimal for each player P_i to choose Z_i. Note that since orders have no impact, players prefer not to give orders. So the unique equilibrium is for no player to give orders, each player P_i to do Z_i, and neither player to quit. This is thus a 'No Authority-Stay' equilibrium.

Consider now the case that $w \in [B + \tilde{\alpha}_1 q - \underline{u}_1, B + \tilde{\alpha}_1 \nu_1 - \underline{u}_1]$, so P_1 quits unless $D_1 = D_2 = Z_1$. It follows that – in any equilibrium where P_1 does not quit for sure – P_1 will always choose $D_1 = Z_1$. So there's two possible type of equilibria: those where P_2 obeys and those where P_2 does as he likes (and P_1 quits if they disagree). Consider first the possibility of a P_1-Au equilibrium. This requires that P_2 always stays and prefers to obey. Always staying requires $w \geq -\tilde{\alpha}_2(1-\nu_2) + \underline{u}_2$. Preferring to obey requires

$$\tilde{\alpha}_2(1-\nu_2) + w \geq p\underline{u}_2 + (1-p)(\tilde{\alpha}_2\nu_2 + w)$$

or

$$w \geq \tilde{\alpha}_2 \left(\frac{(1-p)}{p}(2\nu_2 - 1) - (1-\nu_2) \right) + \underline{u}_2$$

which implies $w \geq -\tilde{\alpha}_2(1-\nu_2) + \underline{u}_2$. So P_1-Au is an equilibrium (and the unique one on the interior) when

$$w \in [\max\left(B + \tilde{\alpha}_1 q - \underline{u}_1, \tilde{\alpha}_2(\kappa(2\nu_2 - 1) - (1-\nu_2)) + \underline{u}_2\right), B + \tilde{\alpha}_1\nu_1 - \underline{u}_1]$$

where $\kappa = \frac{(1-p)}{p}$.

Consider next the possibility of a 'No Authority-Quit' equilibrium, where P_2 chooses Z_2 but P_1 quits when they take different actions. This 'No Authority-Quit' equilibrium requires – beyond the conditions already specified – that P_2 stays upon agreement and prefers not to obey. Staying upon agreement requires

$$w \geq -\tilde{\alpha}_2\nu_2 + \underline{u}_2$$

so that overall, we need

$$w \in [\max(-\tilde{\alpha}_2\nu_2 + \underline{u}_2, B + \tilde{\alpha}_1 q - \underline{u}_1), \min(B + \tilde{\alpha}_1\nu_1 - \underline{u}_1, \tilde{\alpha}_2(\kappa(2\nu_2 - 1) - (1-\nu_2)) + \underline{u}_2)]$$

Analogous results obtain for P_1.

So if $w \in [0, B]$ then we have (only) the following equilibria (where it is not ex-ante known that players will quit) and equilibrium conditions:

1. P_1-Au holds if

$$w \in [\max\left(B + \tilde{\alpha}_1 q - \underline{u}_1, \tilde{\alpha}_2(\kappa(2\nu_2 - 1) - (1-\nu_2)) + \underline{u}_2\right), B + \tilde{\alpha}_1\nu_1 - \underline{u}_1]$$

   The total payoff is then $U_{Au-1} = B + \tilde{\alpha}_1\nu_1 + \tilde{\alpha}_2\frac{1}{2}$

2. P_2-Au holds if

$$w \in [-\tilde{\alpha}_2\nu_2 + \underline{u}_2, \min\left(-\tilde{\alpha}_2 q + \underline{u}_2, B - \tilde{\alpha}_1\left(\kappa(2\nu_1 - 1) - (1-\nu_1)\right) - \underline{u}_1\right)]$$

   The total payoff is then $U_{Au-2} = B + \tilde{\alpha}_2\nu_2 + \tilde{\alpha}_1\frac{1}{2}$

3. NAu-Stay holds if

$$w \in [-\tilde{\alpha}_2 q + \underline{u}_2, B + \tilde{\alpha}_1 q - \underline{u}_1]$$

   The total payoff is then $U_{NAu-Stay} = B + ((1+2\gamma)\eta - 2\epsilon)\frac{\nu_1 + q}{2}$

4. NAu-Quit holds if

$$w \in [\max(-\tilde{\alpha}_2\nu_2 + \underline{u}_2, B + \tilde{\alpha}_1 q - \underline{u}_1), \min(B + \tilde{\alpha}_1\nu_1 - \underline{u}_1, \tilde{\alpha}_2(\kappa(2\nu_2 - 1) - (1-\nu_2)) + \underline{u}_2)]$$

   or if

$$w \in [\max(-\tilde{\alpha}_2\nu_2 + \underline{u}_2, B - \tilde{\alpha}_1\left(\kappa(2\nu_1 - 1) - (1-\nu_1)\right) - \underline{u}_1), \min\left(-\tilde{\alpha}_2 q + \underline{u}_2, B + \tilde{\alpha}_1\nu_1 - \underline{u}_1\right)]$$

   The total payoff is then $U_{NAu-Quit} = \frac{B + ((1+2\gamma)\eta - 2\epsilon)\nu_1}{2} + p\frac{\underline{u}_1 + \underline{u}_2}{2} + (1-p)\frac{B + ((1+2\gamma)\eta - 2\epsilon)q}{2}$

If $w \notin [0, B]$, then w is paid only upon success so that it becomes a share of the residual income. Moreover, at least one player will quit in stage 3b, so that B is lost in that case. It follows that $w \notin [0, B]$ is Pareto-dominated by a contract where w is included in α so that $\tilde{\alpha}_i = \alpha_i + \gamma + v_i$ and $w = 0$.
For later purposes, I will now show that in the absence of w-feasibility constraints, $\hat{U}_{\text{ne}} > \max(\hat{U}_{NAu-Stay}, \hat{U}_{NAu-Quit})$ and (in the case with moral hazard, i.e, assumption 1b) $\hat{U}_{\text{e}} > \hat{U}_{\text{ne}}$ where $\hat{U}_{\text{e}}$ and $\hat{U}_{\text{ne}}$ are the maximal total utility of an authority equilibrium respectively with and without effort and $\hat{U}_{NAu}$ is the maximal total utility of the respective no-authority equilibrium. Note that if $\hat{U}_{\text{ne}} > \max(\hat{U}_{NAu-Stay}, \hat{U}_{NAu-Quit})$ under assumption 1b then it will also hold under 1a, so it suffices to consider only assumption 1b.
Consider then first NAu. Since the allocation of residual income does not affect $\hat{U}_{NAu}$ (except through its impact on effort), effort is efficient, and $\frac{e}{(1-\theta)} \leq \frac{1}{2}$, there will always be effort in equilibrium. For NAu-Quit, it follows that

$$\begin{aligned}
\hat{U}_{\text{ne}} - \hat{U}_{NAu-Quit} &= B + (1+\gamma)\theta\nu_1 + \gamma\theta\frac{1}{2} - \frac{B + ((1+2\gamma) - 2e)\nu_1}{2} - \frac{\underline{u}}{2} \\
&= \frac{B + \nu_1 - \underline{u}}{2} + (1+\gamma)\theta\nu_1 + \frac{\gamma\theta}{2} - (1+\gamma-e)\nu_1 + (1-p)\frac{(\underline{u} - B - q) - 2\gamma q + 2eq}{2} \\
&\geq \frac{\epsilon}{2} - (1-\theta)\nu_1 + \frac{\gamma\theta}{2} + (e - (1-\theta)\gamma)\nu_1 + (1-p)\frac{\epsilon - 2\gamma q}{2}
\end{aligned}$$

Assumption 1b implies that $(1-\theta) \leq \frac{\epsilon}{2}$ and $\gamma \leq \frac{e}{(1-\theta)}$ so that indeed $\hat{U}_{\text{ne}} > \hat{U}_{NAu-Quit}$.
For NAu-Stay, finally, it follows that

$$\begin{aligned}
\hat{U}_{\text{ne}} - \hat{U}_{NAu-Stay} &= B + (1+\gamma)\theta\nu_1 + \gamma\theta\frac{1}{2} - B - ((1+2\gamma) - 2e)\frac{q+\nu_1}{2} \\
&= \theta\nu_1 - \nu_1 + \frac{\nu_1 - q}{2} + \gamma\left(\theta\frac{1}{2} + \theta\nu_1 - q - \nu_1\right) + e(q+\nu_1) \\
&\geq \frac{2\epsilon}{2} - (1-\theta)\nu_1 - 2\gamma
\end{aligned}$$

where I use that – by assumption 1b – $\nu_1 - q = B + \nu_1 - \underline{u} + \underline{u} - B - q \geq 2\epsilon$. Assumption 1b implies that $(1-\theta) \leq \frac{\epsilon}{2}$ and $\gamma \leq \frac{\epsilon}{4}$ so that indeed $\hat{U}_{\text{ne}} > \hat{U}_{NAu-Stay}$.
Consider finally Au-P_1 under assumption 1b. (The analysis and result for Au-P_2 is completely analogous.) Since moving residual income from P_2 to P_1 increases $\hat{U}$, either $\alpha_1 = 1$ (with no effort) or $\alpha_1 = 1 - \left(\frac{e}{(1-\theta)} - \gamma\right)$ (with effort by both players). The difference in utility is then

$$\begin{aligned}
\hat{U}_{\text{e}} - \hat{U}_{\text{ne}} &= B + (1 + 2\gamma - \frac{e}{(1-\theta)} - e)\nu_1 + \left(\frac{e}{(1-\theta)} - e\right)\frac{1}{2} - B - (1+\gamma)\theta\nu_1 - \gamma\theta\frac{1}{2} \\
&= \left((1-\theta)(1+\gamma) + \gamma - \frac{e}{(1-\theta)} - e\right)\nu_1 + \left(\frac{\theta e}{(1-\theta)} - \theta\gamma\right)\frac{1}{2}
\end{aligned}$$

Assumption 1b implies that $(1-\theta) \geq \frac{e}{(1-\theta)} + e$ and $\gamma < \frac{e}{(1-\theta)}$ so that indeed $\hat{U}_{\text{e}} > \hat{U}_{\text{ne}}$.
This finalizes the proposition. ■

References

Aghion, P., M. Dewatripont, and P. Rey (2004): "Transferable Control," *Journal of the European Economic Association*, 2(1), 115– 138, Working Paper.

Aghion, P., and J. Tirole (1997): "Formal and real authority in organizations," *Journal of Political Economy*, 105(1), 1– 29.

Alchian, A. A., and H. Demsetz (1972): "Production, Information Costs, and Economic Organization," *American Economic Review*, 62, 777– 795.

Arrow, K. J. (1974): *The Limits of Organization*. W. W. Norton & Co, New York.

Baker, G., R. Gibbons, and K. J. Murphy (2006): "Contracting for Control," Working Paper.

Baker, G. P., R. S. Gibbons, and K. J. Murphy (1999): "Informal Authority in Organizations," *Journal of Law, Economics, and Organization*, 15(1), 56– 73.

Barnard, C. (1938): *The Functions of the Executive*. Harvard University Press, Cambridge MA.

Boot, A. W. A., R. Gopalan, and A. V. Thakor (2006): "The Entrepreneurs Choice Between Private and Public Ownership," *Journal of Finance*, 61(2), 803– 836.

Coase, R. H. (1937): "The Nature of the Firm," *Economica*, 4, 386– 405.

Dittmar, A., and A. V. Thakor (2006): "Why Do Firms Issue Equity?," *Journal of Finance*, Forthcoming.

Fayol, H. (1916): "Administration Industrielle et Generale," *Bulletin de la Societe de l'Industrie Minerale*, 10, 5– 164.

——— (1949): *General and Industrial Management*. Sir Isaac Pitman, London.

Gans, J. (2005): "Markets for Ownership," *Rand Journal of Economics*, 36(2), 433– 455.

Gibbons, R. (2005): "Four Formal(izable) Theories of the Firm?," *Journal of Economic Behavior and Organization*, 58, 200– 245.

Grossman, S. J., and O. D. Hart (1986): "The costs and benefits of ownership : A theory of vertical and lateral integration," *Journal of Political Economy*, 94(4), 691– 719.

Gul, F. (1989): "Bargaining Foundations of Shapley Value," *Econometrica*, 57(1), 81– 95.

Harrison, M., and D. M. Kreps (1978): "Speculative Investor Behavior in a Stock Market with Heterogenous Expectations," *Quarterly Journal of Economics*, 92(2), 323– 336.

Harsanyi, J. C. (1968): "Games with Incomplete Information Played by 'Bayesian' Players, I-III, Part III. The Basic Probability Distribution of the Game," *Management Science*, 14(7), 486– 502.

Hart, O., and B. Holmstrom (2002): "A Theory of Firm Scope," Working Paper.

Hart, O., and J. Moore (2006): "Contracts as Reference Points," Working Paper.

Hart, O. D. (1995): *Firms, contracts, and financial structure*. Clarendon Press, Oxford.

Hart, O. D., and J. Moore (1990): "Property rights and the nature of the firm," *Journal of Political Economy*, 98(6), 1119– 1158.

Hermalin, B. E. (1999): "The Firm as a Noneconomy: Some Comments on Holmstrom," *Journal of Law, Economics, and Organization*, 15(1), 103– 106.

Holmstrom, B. (1999): "The Firm as a Subeconomy," *Journal of Law, Economics, and Organization*, 15(1), 74– 102.

Holmstrom, B., and P. Milgrom (1991): "Multi-Task Principal Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization*, 7, 24– 52.

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

Holmstrom, B., and J. Roberts (1998): "The Boundaries of the Firm Revisited," *Journal of Economic Perspectives*, 12(4), 73– 94.

Kahn, C., and G. Huberman (1988): "Two-Sided Uncertainty and 'Up-or-out' Contracts," *Journal of Labor Economics*, 6(4), 423– 444.

Klein, B., and K. B. Leffler (1981): "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy*, 89(4), 615– 641.

Klepper, S., and P. Thompson (2006): "Intra-industry Spinoffs," Working Paper.

Knight, F. (1921): *Risk, Uncertainty and Profit*. Houghton-Miller, Boston and New York.

Levin, J., and S. Tadelis (2004): "Employment versus Contracting in Procurement: Theory and Evidence from U.S. Cities," Mimeo, Stanford University.

Marino, A. M., J. G. Matsusaka, and J. Zábojník (2006): "Authority and Disobedience," .

Marschak, J., and R. Radner (1972): *Economic Theory of Teams*. Yale University Press, New Haven, CT.

Milgrom, P., and J. Roberts (1988): "An Economic Approach to Influence Activities in Organizations," *The American Journal of Sociology; Supplement: Organizations and Insitutions*, 94, 154– 179.

——— (1990): "The economics of modern manufacturing : Technology, strategy, and organization," *American Economic Review*, 80(3), 511– 528.

Monteverde, K., and D. J. Teece (1982): "Supplier Switching Costs and Vertical Integration in the Automobile Industry," *Bell Journal of Economics*, 25(4), 206– 213.

Morris, S. (1994): "Trade with Heterogeneous Prior Beliefs and Asymmetric Information," *Econometrica*, 62(6), 1327– 1347.

——— (1995): "The Common Prior Assumption in Economic Theory," *Economics and Philosophy*, 11, 227– 253.

Prendergast, C. J. (1995): "A Theory of Responsibility in Organizations," *Journal of Labor Economics*, 13(3), 387– 400.

Rajan, R. G., and L. Zingales (1998): "Power in a Theory of the Firm," *Quarterly Journal of Economics*, 113, 387– 432.

Simon, H. (1951): "A Formal Theory of the Employment Relationship," *Econometrica*, 19, 293– 305.

Van den Steen, E. J. (2001): "Essays on the Managerial Implications of Differing Priors," PhD Dissertation, Stanford University.

——— (2005a): "On the Origin of Shared Beliefs (and Corporate Culture)," MIT Sloan Research Paper No. 4553-05.

——— (2005b): "Too Motivated?," MIT Sloan Working Paper No. 4547-05.

——— (2006a): "Disagreement and the Allocation of Control," Working Paper MIT Sloan.

——— (2006b): "An Economic Theory of the Firm as a Legal Entity," Working Paper.

——— (2006c): "The Limits of Authority: Motivation versus Coordination," Mimeo MIT-Sloan.

Weitzman, M. L. (1974): "Prices vs. Quantities," *Review of Economic Studies*, 41(4), 477– 491.

Wernerfelt, B. (1997): "On the Nature and Scope of the Firm: An Adjustment-Cost Theory," *The Journal of Business*, 70(4), 489– 514.

——— (2002): "Why Should the Boss Own the Assets?," *Journal of Economics and Management Strategy*, 11(3), 473– 485.

Williamson, O. (1975): *Markets and Hierarchies : Analysis and Antitrust Implications*. Free Press, New York.

Yildiz, M. (2003): "Bargaining Without a Common Prior - An Immediate Agreement Theorem," *Econometrica*, 71(3), 793– 811.