Transferable control

Philippe Aghion, Mathias Dewatripont and Patrick Rey[1]

February 2002

[1]Respectively Harvard University, University College London and CEPR; ECARES (Universite Libre de Bruxelles) and CEPR; and Université de Toulouse and CEPR. We thank Oliver Hart, Jean Tirole, seminar participants in the International Francqui Conference (Brussels, November 1999) on the Economics of Contracting, Mark Armstrong and two referees for useful comments and suggestions.

Abstract

This paper investigates contracting situations with *transferable control*: control cannot be allocated contractually but can be transferred, provided it is in the interest of the individual who transfers it. We illustrate the relevance of this notion in a context with incomplete information in which transferring control over particular actions to another party allows that party to build a reputation regarding her willingness to cooperate in the future. We also relate the concept of transferable control to the notions of formal and real authority, developed by Aghion-Tirole (1997): control is transferable when for example formal authority can be contracted upon but the information needed to "really" exercise authority cannot; in such situation, control can indeed be credibly transferred – by providing the relevant information – but has to be incentive-compatible.

1 Introduction

Much progress has been accomplished in the last fifteen years in modelling control allocation and in using this notion to analyze vertical and lateral integration,[1] financing decisions[2] and the allocation of authority within firms.[3] In all these models, although actions may not be contractible (either ex ante, or both ex ante and ex post), the allocation of control is. In this paper, we argue that contracting on control may in fact be problematic. To understand this, consider for example the distinction between formal and real authority introduced by Aghion and Tirole (1997): the actual exercise of authority requires information absent which formal control rights can be vacuous. Now, while it is natural to assume that formal control rights can be contracted upon, the same is not true about information transfers. In this context, we are led to define the concept of *transferable control*, a situation where the party initially in control can transfer this control to another party but cannot commit herself to do so – if for example some information is required, the party may claim that she has no useful information, or provide useless information; she will thus transfer control only she finds it in her interest to do so.

In this paper, we argue that the concept of transferable control provides useful insights. We consider the following contracting situation between a principal (P) and an agent (A): (i) a first action (project design), over which control is either fully contractible or simply transferable from P to A, is followed by a second action (project implementation) by P; (ii) agent A has private information about his willingness to act cooperatively, which affects the outcome of the project. The optimal contract depends upon whether control over the first action is contractible or only transferable: in the former case, message games result in transferring control to A precisely when he is a non-cooperative type; in the latter case, message games are not used and instead P transfers control to A in order to allow him to build a reputation regarding his willingness to cooperate. This second case is consistent for example with the widespread practice of writing initial short-term contracts in order to "test" a new partner, before entering into

[1]See for example Grossman and Hart (1986) Hart and Moore (1990) and more generally Hart (1995).

[2]See for example Aghion and Bolton (1992), Hart and Moore (1994), Dewatripont and Tirole (1994).

[3]See for example Aghion and Tirole (1997) and Dessein (2000, 2002).

a more sustained relationship. Banerjee and Duflo's (1999) study of the Indian customized software industry provides an illustration of such two-step contracting process: when a big US computer firm starts contracting with a small Indian software firm without reputation, it signs a contract which allows the small firm to build a reputation, while nonetheless imposing a significant share of the risk on the small firm. Once previous contracting relationships have enhanced the reputation of the small firm, the parties then move to contracts where the big firm takes more of the risk. Control transfers to test other parties' ability or willingness to cooperate are also commonplace within firms and organizations, especially when it comes to delegate authority to subordinates. The analysis thus illustrates two related insights: first, the power of message games recedes as one moves from contractible to transferable control actions; second, in situations where future cooperation is uncertain, optimal contracts are more realistic if control is transferable rather than contractible.

To position the notion of transferable control within the contract theory literature, it is convenient to refer to the degree of contractibility of actions. More specifically, consider the following polar cases:

(i) at one extreme, a world with fully contractible actions: in such a world, the relationship is fully determined by the initial contract or mechanism; this case encompasses the *"classical" implementation literature* à la Maskin (1999) or Moore-Repullo (1988), where one can contract on entire game forms; in this case, the contracting parties can limit themselves to sending (possibly sequential) messages to a "planner" who then takes or dictates all relevant actions;[4]

(ii) at the other extreme, a world with only noncontractible and pre-assigned actions: then, the structure of the game played by the parties is hardly affected by the initial contract; this case encompasses *standard game-theoretic models*, where there is no room for contracting at all, but also *moral hazard models* (à la Mirrlees (1999), Holmström (1979, 1982), Legros-Matthews (1993), ...), where actions are noncontractible and pre-assigned, but contracts over output variables, correlated with the actions, can influence the choice of actions.

In-between, there is *partial contracting* (see Aghion et al. (2002)), which we

[4]The implementation results of this literature have been generalized by Maskin-Tirole (1999a) to the case where actions are noncontractible "ex ante", before the revelation of the state of nature, but become contractible "ex post", after the revelation of the state of nature.

define as situations where formal contracts, rather than determining the entire relation between the contracting parties, can only influence the *underlying game* between them.[5] These are situations in which some actions are nonverifiable and therefore not contractible, so that those actions cannot be delegated to – or dictated by – a social planner. Yet *control* over such actions may be fully or partly contractible, so that the contract can specify (initial) control rights over these actions, which in turn influences the dynamic relationship between the parties. Much of the existing literature on control rights and authority, has concentrated on partial contracting situations where control is fully contractible.[6] The main focus of this paper is on the case, intermediate between the case of contractible control and that of pre-assigned actions (or "moral hazard"), where control over particular actions is not contractible but still credibly transferable.

Most closely related to this paper is the above mentioned literature on control rights and authority, in particular Aghion-Tirole (1997), Dessein (2002), and Hart-Moore (1999b). We contribute to this literature by allowing for the possibility of credible control transfers.[7] Our model is also related to the game-theoretic literature on reputation, e.g. Kreps et al. (1982), Sobel (1985) and Watson (1999) for example. We differ from these papers by adding a contracting stage to the game theoretic approach, and by analyzing under which conditions – in particular on the extent to which control is contractible versus transferable – reputational equilibria, involving delegation and trust building, can actually exist. In particular, we show that trust building is dominated by direct revelation mechanisms when actions and/or control over these actions are contractible. Focusing on control allocation as a way to induce cooperation, our analysis is also related to the literature on "formal" versus "informal" contracting, and in particular with Baker et al.(1997) and Halonen (1997). In a repeated model of ownership allocation à la Grossman-Hart (1986), where all parties have complete information, Baker et al. (1997) show that vertical integration may help the parties to hold

[5] In Aghion et al (2002) we discuss the connection between partial contracting and the debate on the foundations of incomplete contracts (e.g in Segal (1999), Hart-Moore (1999a), Maskin-Tirole (1999a) and Tirole (1999).

[6] E.g, see Aghion-Bolton (1992), Dewatripont-Tirole (1994), Aghion-Tirole (1997), Legros-Newman (1999) and Hart-Moore (1999b).

[7] In Aghion-Tirole (1997) the party that has acquired the relevant information about which project to implement, always keeps that information and always proposes her prefered project. She never transfers that information to the other party. Thus real authority is never transferred in that model.

on to their promises of taking costly actions or making costly monetary transfers. Based also on a repeated ownership allocation model, but with imperfect information about the parties' disutility from cheating on effort or investment commitments, Halonen (1997) argues that joint ownership may emerge as a desirable contractual outcome ex ante, because both parties then find it particularly costly reputationwise to renege on their promises. This rationale for control allocation is reminiscent of the argument of Boot et al. (1993), who stress that "loan commitment contracts" that allow banks to unilaterally renege on their commitments imply that the bank's reputation for "fairness" is enhanced when they do not actually renege. These papers, however, do not distinguish between contractible control and transferable control but instead restrict attention to simple contracts, even though revelation mechanisms would be more effective (for example in Halonen's case where trade and cheating are both supposed to be ex post verifiable).

The paper is organized as follows. Section 2 describes the framework. Section 3 focuses on contractible control and shows that the optimal contract is a revelation mechanism promising control to the agent when he announces a non-cooperative type. Section 4 studies the situation where control is not contractible but only transferable; it can then become optimal to delegate control, in order to "test" the agent. In section 5 we analyze the robustness of our conclusions to the introduction of monetary transfers as a potential screening variable. Section 6 provides a foundation for our notion of transferable control by connecting it to the concepts of formal and real authority of Aghion and Tirole (1997); section 7 uses this foundation to analyze the scope of delegation, in an extension of the basic framework in which control can be partly transferred.

2 Framework

This section outlines an incomplete information framework where control allocation serves as a natural instrument to induce "reputation building", thereby enhancing trust and cooperation in the relation between the contracting parties. Specifically, we consider a principal-agent (hereafter P and A) relationship meant to carry out a project. This project involves two stages, design and implementation, preceded by a contracting phase. In the design stage, the party in control chooses between two actions, C and N; action C is the "cooperative" action that is best for the project, whereas action N is a "non-cooperative" action that the agent may favor – for example, it may enhance the agent's general human capital or, more generally, his market value. In the implementation stage, P decides whether to "cooperate," i.e. implement the project (I), or to stop it (S). The principal initially has overall control over the project but needs the agent to implement it. A can be "good" or "bad," and the project is worth implementing only if A is good. Initially, P does not know A's type; we denote by μ her prior probability that A is bad.

Specifically:

- in stage 1, the design decision has to be made; P can either take the decision or let A take it (we shall distinguish between the case where control over project design is *contractible* and the case where it is simply *transferable* by the principal); the decision itself is observed by both parties but not contractible.
- in stage 2, P freely decides whether to implement the project: his decision is not contractible and cannot be delegated to A; if implemented, the outcome of the project depends upon A's type.

Since the focus is on how contracts can affect this game, for the sake of presentation we will adopt a simple payoff structure; for each pair of stage 1 and stage 2 actions, the payoffs of the principal and of the agent (in that order) are given by the following tables:

- when A has a good type:

Action	I	S
C	$1, 1$	$0, 0$
N	$1 - l, 1$	$-l, 0$

- when A has a bad type:

Action	I	S
C	$-L, b$	$0, 0$
N	$-L-l, B+b$	$-l, B$

Figure 1

where:

$$B > b > 0,$$
$$L > l > 0.$$

That is, a good A is willing to cooperate in stage 1 and gains 1 from P's implementing the project; in contrast, a bad A gains B from the non-cooperative design action N; he also gains b from the implementation of the project, but prefers the non-cooperative design action N even if this induces P to stop the project ($B > b$).[8] P incurs a loss l from the non-cooperative action at the design stage and an even bigger loss L from implementing the project when the agent has a bad type; P is thus willing to let a bad A choose the non-cooperative action at the design stage to learn his type (and stop the project).

If P is uninformed about A's type when deciding whether to implement or stop the project in stage 2, she will stop the project whenever:

$$(1-\mu) + \mu(-L) < 0,$$

or equivalently:

$$\mu > \mu^* \equiv \frac{1}{1+L}.$$

Incomplete information thus generates two types of problems: first, when μ is too large, P prefers to stop the project since she cannot obtain a positive payoff in stage 2. Second, when μ is small enough, P cooperates (does not stop the

[8]Fixing the implementation decision, a good type is here indifferent between cooperating or not on design. The analysis applies unchanged when a good A gains ε from the non-cooperative action N, where ε is small but either negative (a good type strictly prefers to cooperate) or positive (a good type is slightly reluctant to cooperate). What matters is that a good type is willing to cooperate if this induces P to implement the project (i.e., $1 > \varepsilon$).

Likewise, the analysis could be extended to allow the bad type's benefits from I or N to be non-additive.

project) but loses L when A is a bad type. We now explore alternative means whereby these two problems can be solved.

Except in Section 5, we assume that payoffs are private benefits and the parties are not responsive to monetary incentives. Therefore, contracts will consist of revelation mechanisms to be played at the beginning of each stage; as a function of messages sent at the beginning of stage 1, control over project design is allocated to the agent or kept by the principal, and given the messages exchanged between the two parties up to stage 2, the principal decides whether to implement the project.

3 Contractible control

In this section we assume that control over stage 1 can be specified by an enforceable contract between P and A. The set of feasible strategies and contracts and the timing of moves can then be described as follows.

In the contracting phase, P offers a contract to A; this contract dictates an allocation of control over the stage 1 action, possibly contingent upon messages sent by A at that stage. The agent then decides whether or not to accept this contract; if he refuses, the game ends and both parties get zero; if he accepts, the game proceeds as follows:

- In stage 1, A sends messages and control is then allocated to P or A according to the contract; whoever ends up in charge of stage 1 chooses between C and N.
- In stage 2, A may again send messages; P then decides whether or not to implement the project.

Note first that, without loss of generality, one can restrict attention to contracts in which the agent sends no message in stage 2. This follows from the cheap talk nature of the stage 2 message game: a bad A only sees advantages and no cost in mimicking a good type at stage 2, since doing so can only encourage P to implement the project, which is good for him. And since P's decision to implement the project or not is not contractible, she will do it if and only if it is in her interest to do so.

In this paper, we shall allow for random control allocations but, for simplicity, we shall concentrate on contracts where pure strategies are played. Two types of contracts are therefore possible:

- Contracts where the agent does not send any message (equivalently, both types send the same message which is then useless). P can for example simply choose to keep control of the stage 1 action, and implement the project in stage 2 if and only if $\mu < \mu^*$. P's payoff is then $\max\{(1-\mu)-\mu L, 0\}$. Alternatively, P could give control to A and infer A's type from his choice of action; we further explore this latter possibility in the following section.
- *Revelation mechanisms*, where P learns A's type through the messages sent in stage 1 and then implements the project only if A reports a good type.

Intuitively, in such a revelation mechanism a bad A could gain b in stage 2 by misreporting a good type. To prevent this, P must reward a bad A for revealing himself, and can do so by giving a bad A control over stage 1. Since a bad A can obtain B by choosing N, to induce truthtelling P must grant control to a bad A with at least probability b/B; in addition, in order to minimize a bad A's incentive to misreport his type P should not give control to a good A. The following proposition confirms this intuition:

Proposition 1 *P's optimal revelation mechanism, M_c, is such that: (i) P keeps control in stage 1 when A announces a good type, and (ii) P allocates control to A with probability b/B when A announces a bad type. In the associated equilibrium, action N is chosen if and only if a bad A gets control and the project is implemented if and only if A announces a good type.*

Proof. In any revealing equilibrium, P implements the project if the agent reports a good type and stops it otherwise. Thus, a good A is always willing to report his type, since he gets 1 by doing so and 0 by behaving as a bad type.

Let x and y denote the probabilities that A obtains control when announcing a good and a bad type. P will clearly choose action C when she has control over stage 1. When a good A gets control, he is indifferent between actions C and N while P benefits from action C. In contrast, a bad A always chooses action N when in control. Therefore, the best for P is that a good A chooses C whenever in control: this improves P's payoff, and also helps P deterring a bad A from misreporting a good type. P's expected payoff is then given by

$$1 - \mu - \mu y l,$$

while the bad A's incentive compatibility condition for truthtelling is:

$$yB \geq (1 - x) b + xB.$$

Since $B > b$, the optimal probabilities of control are thus $x = 0$ and $y = b/B$. ■

The mechanism M_c gives P an expected payoff equal to

$$1 - \mu - \mu \frac{b}{B} l,$$

and is thus positive whenever

$$\mu < \mu_c \equiv \frac{1}{1 + \frac{b}{B} l}.$$

The mechanism M_c addresses the two problems mentioned above: the revelation of A's type allows P to implement the project when – and only when – it is desirable to do so. However, this revelation has a cost: P must "reward" a bad type A for telling the truth, namely by granting control to that bad type in stage 1 with probability b/B, in which case action N is implemented instead of action C.

Note that P prefers M_c over "keeping control with probability 1 and implementing the project": $L > l$ and $B > b$ imply

$$1 - \mu - \mu \frac{b}{B} l > 1 - \mu - \mu L.$$

In addition, $\mu_c < \mu^*$, so there are cases ($\mu_c < \mu < \mu^*$) where M_c is profitable even when, if uninformed, P would choose to stop the project in stage 2.

4 Transferable control and reputation building

While M_c illustrates how control allocation can be used to induce truth-telling through a standard revelation mechanism, this contract suffers from obvious credibility problems: P has no incentives to transfer control to a bad A once his type has been revealed. Hence our interest in exploring the case where control is transferable but noncontractible: in that case, P can *choose* to transfer control (given A's messages) at the beginning of stage 1, but cannot *commit* to do so at the contracting stage.

In that case, the set of strategies and the timing of events are modified as follows. In the contracting phase, P offers a contract to A, which again allows for messages to be sent by A at the beginning of each stage, but can no longer dictate the allocation of control over the first stage. As before, A then decides whether or not to accept the contract; if he refuses, the game ends and both parties get zero; if he accepts, the game proceeds as follows.

- In stage 1, A sends messages; then P decides *with full discretion* whether or not to transfer control to A; whoever ends up in charge of stage 1 chooses between C and N.
- Stage 2 is unchanged: A may again send messages before P decides whether or not to implement the project.

We can immediately establish the following lemma:

Lemma 2 *When control over stage 1 is transferable but not contractible, there is no loss of generality in not asking the agent to send messages.*

Proof. When control over stage 1 is not contractible, message games involve cheap talk not only in stage 2 but also in stage 1: in both stages now, a bad A sees only advantages and no cost in reporting a good type. Therefore the principal may as well ignore any message A might send at any stage. ∎

Given this lemma, at the contracting stage P must simply choose between keeping control, in which case she remains uninformed about A's type by the end of stage 1, or transferring stage-1 control to A. The cost of such a control transfer is that the bad A will choose action N, which is his dominant strategy. The good A instead is happy to choose action C, especially since this signals her good type and ensures that P implements the project.[9] In comparison with the case where

[9] We disregard here the "bad" equilibrium where the good A also chooses action N, because action C would lead P not to implement the project: such an equilibrium does not survive the Cho-Kreps "intuitive criterion", since action C is dominated for the bad A (but not for the good A).

P keeps control in stage 1, transferring control to A clearly improves A's expected payoff. It also increases P's expected payoff whenever the (short-term) loss from losing control in stage 1 is more than compensated by the (long-term) benefit of learning A's type prior to stage 2.

The stage-1 loss comes from the fact that a bad A chooses action N. The expected loss for P, who then stops the project in stage 2, is therefore equal to μl. The stage-2 informational gain now depends upon the equilibrium that would prevail if A did not signal his type:

- If the probability of a bad type agent is sufficiently small ($\mu \leq \mu^*$), an uninformed principal P would always implement the project; learning A's type then allows P to stop the project and avoid the loss L when A turns out to be a bad type. Thus, for $\mu \leq \mu^*$ learning A's type allows P to save μL in expected terms; P will thus prefer to grant control to A in stage 1 since $L > l$ implies

$$\mu L \geq \mu l.$$

- If the probability of a bad type agent is sufficiently high ($\mu > \mu^*$), an uninformed principal P would instead stop the project; learning A's type then allows P to implement the project and gain 1 if A turns out to be a good type. Thus, learning A's type when $\mu > \mu^*$ generates an additional expected gain equal to $(1 - \mu)$ to the principal; for such values of μ, P will thus prefer to grant control to A in stage 1 if:

$$1 - \mu > \mu l,$$

or equivalently:

$$\mu < \mu_t \equiv \frac{1}{1+l}.$$

This establishes the following:

Proposition 3 *P's optimal transferable-control contract, M_t, is such that: (i) no messages are sent before stage* 1 *(and a fortiori after that stage); (ii) P transfers control over the stage 1 action to A if $\mu < \mu_t$ and keeps control (and stops the project in stage 2) otherwise; (iii) if A obtains control in stage 1, he chooses action C if his type is good and action N otherwise, and P implements the project if and only if action C has been chosen.*

It is therefore in P's interest to transfer control to A for small values of μ: if μ is too high, the hope that a good A will act cooperatively in stage 1 in order to keep the project going in stage 2 is too small, compared with the short-term loss from a bad A's non-cooperating in the first stage.

To summarize, when A' willingness to cooperate is initially unknown by P, two problems may potentially arise: either cooperation is impossible in stage 2 (P stops the project), or "excessive" cooperation imposes losses on P. Then, "testing" A by giving him control over stage 1 creates an opportunity for A to build up a reputation for trustworthiness, which in turn helps overcome each of the above two problems. By transferring control of stage 1 to A, P may not lose that much since a good A will want to choose action C, especially since this induces the implementation of the project; furthermore, any early loss induces P to take "appropriate measures" (that is, to stop the project) to prevent subsequent losses.

A final remark to conclude this section: by transferring control to A, which is optimal when $\mu < \mu_t$, P achieves an expected payoff equal to:

$$1 - \mu - \mu l.$$

This payoff achieved through M_t is lower than what she gets with the contractible-control revelation mechanism M_c described in the previous section, namely:

$$1 - \mu - \frac{b}{B}\mu l.$$

Therefore:[10]

- When control is contractible, it is optimal for P to have A's type revealed through a direct revelation mechanism rather than through trust-building; specifically, the relationship is profitable when $\mu < \mu_c$ and M_c is then the optimal contract for P, since it induces truth-telling with a smaller probability (b/B instead of 1) of control allocated to a bad A.

[10] Specifically, when P can contract over stage 1 control, the relationship is profitable if and only if $\mu < \mu_c$, and M_c is then the optimal contract for P. When control is transferable but not contractible, the relationship is profitable only when $\mu < \mu_t < \mu_t$ and M_t is then the optimal contract for P.

- When control is transferable but not contractible, the relationship is profitable only when

$$\mu < \mu_t = \frac{1}{1+l},$$

and M_t is then the optimal contract for P; control is then transferred with higher probability (1 instead of b/B) but only for a smaller interval of μ's (since $\mu_t < \mu_c$).

5 Monetary responsiveness

We have so far restricted attention to the case where the contracting parties do not respond to monetary incentives. But now suppose that P's and A's utilities are given by:

$$U_P = \pi_P + p \quad \text{and} \quad U_A = \pi_A - p$$

where π_P and π_A denote the private benefits of the two parties, and p is a monetary transfer from A to P.

Allowing for monetary transfers makes the two parties' utilities transferable. Therefore, what matters for efficiency and welfare is the sum of both parties' payoffs; in keeping with the analysis of the previous sections, we assume in this section

$$(L >)\, l > B\, (> b)\,,$$

which implies that it is efficient to stop the project when the agent is bad.

Allowing for transfers also provides additional ways to acquire the information over A's type. In particular, if $b < 1$ P can simply keep control and obtain full revelation at stage 2 by having the good A pay 1 against project implementation.[11] The possibility of monetary transfers thus eliminates the role for allocating or transferring control to A in that case. When $b > 1$, however, the previous insights remain valid: monetary transfers reduce but do not eliminate the use of control allocation when control is contractible, and transferring control

[11] If $b < 1$, in the absense of any messages there exists a separating equilibrium in which: (i) P keeps control and chooses C in stage 1 and implements the project in stage 2 only if A pays $p = 1$; (ii) only a good A makes the payment. In this equilibrium, P implements the project only when A is good and both types of agent get zero rent.

remains a useful reputation-building device when control in transferable but not contractible.

5.1 Contractible control case

Allowing for monetary transfers enriches the set of strategies and contracts as follows. In the contracting phase, P offers a contract which again allows for messages sent by A at the beginning of each stage, and allocates control on the basis of the messages sent in stage 1; but in addition the contract can now require transfers contingent on all messages. As before, A then decides whether or not to accept the contract; if he refuses, the game ends and both parties get zero; if he accepts, the game proceeds as follows:

- In stage 1, A sends messages and control is allocated accordingly; whoever ends up in charge of stage 1 chooses between C or N.

- In stage 2, A again sends messages and makes monetary transfers, contingent on all messages sent in stages 1 and 2.

Introducing monetary transfers allows P to extract rents from A and also to "sell" stage 1 control. The following proposition extends the previous analysis:

Proposition 4 *Assume that $b > 1$. Then, restricting attention to pure strategies, the optimal separating contract for P consists in selling control at price 1 with probability $(b-1)/(B-1)$ when A reports a bad type, while keeping control and requiring a payment of 1 when A reports a good type. In this mechanism, a good A obtains no rents while a bad A obtains $b-1$ and P obtains a positive expected payoffs as long as*

$$\mu < \mu_c^m = \frac{1}{1 + \frac{b-1}{B-1}\frac{l-1}{2}}.$$

Proof. As in section 3, we can restrict attention to direct mechanisms where, at each stage, A sends at most two messages – "good" and "bad". We proceed again by backward induction.

We first check that there is no scope for revelation of A's type in stage 2. Consider a candidate equilibrium where A reveals his type at stage 2, in which case P implements the project only if A is good. In order for a bad A not to pretend to be good, announcing a good type must involve an additional transfer p in P's favor. Whatever the stage 1 action (C or N), a good A would find it profitable to pay p (and have the project implemented) only if

$$1 - p > 0,$$

and a bad A will not find it profitable to mimic a good type if

$$b - p < 0.$$

But these two conditions cannot be simultaneously satisfied when $b > 1$.

Moving back one step, consider the control allocation stage. A can in principle reveal his type by "acquiring control" at some specified price. More precisely, consider a separating equilibrium and let (x_g, p_g) and (x_b, p_b) denote the agent's probabilities of obtaining control and the payments in P's favor respectively attached to reporting a good type and a bad type. In any such separating equilibrium, when P keeps control of stage 1 she chooses C and then implements the project only if A reported a good type. A bad A chooses N whenever in control, and P then stops the project. Without loss of generality, a good A can be assumed to choose C when in control,[12] and P then implements the project.

The two participation constraints are thus:

$$1 - p_g \geq 0$$

and:

$$x_b B - p_b \geq 0.$$

If a bad A falsely reports a good type, he reveals his type when in control (by choosing N), in which case P then stops the project, and induces P to implement the project when P keeps control. A bad A is thus willing to report the truth if:

$$x_b B - p_b \geq (1 - x_g)\, b + x_g B - p_g.$$

If a good A falsely announces a bad type, P chooses C when in control and then stops project. When A gets control, choosing N would again lead P to stop the

[12] As before, requiring a good A to choose N is inefficient and gives a bad A additional incentives to falsely report a good type.

project; however, at this point A may try to "signal" his good type by choosing C instead. Indeed, following the Cho-Kreps intuitive criterion, a choice of action C can only be attributed to a good A, and must therefore be followed by project implementation, irrespective of previous announcements. The good A's incentive constraint is then:[13]

$$1 - p_g \geq x_b - p_b.$$

The optimal contract leading to a separating equilibrium maximizes P's expected payoff:

$$(1-\mu)(p_g + 1) + \mu [p_b + (1 - x_b) \times 0 - x_b l],$$

subject to the above constraints. Two things can be noted at this point. First, giving control less often to a good type (reducing x_g) relaxes the bad A's incentive constraint (since $B > b$) without affecting P's payoff, so that it is optimal for P to set $x_g = 0$. Second, since $B \geq b > 1$, the participation constraint of the bad A is satisfied whenever his incentive constraint and the participation constraint of the good A are. These two observations allow us to rewrite the relevant constraints as:

$$\begin{aligned} 1 - p_g &\geq 0, \\ 1 - p_g &\geq x_b - p_b, \end{aligned}$$

and:

$$x_b B - p_b \geq b - p_g.$$

The last condition must be binding, otherwise P could increase p_b. Using this condition to determine p_b, P's expected payoff can be expressed as

$$p_g + (1-\mu) + \mu (x_b B - 1),$$

while the first two conditions become

$$\begin{aligned} p_g &\leq 1, \\ x_b (B - 1) &\leq b - 1. \end{aligned}$$

[13] Absent the Cho-Kreps criterion, P may refuse A's signal and stick to the reported type, which would reduce A's benefit from falsely reporting a bad type. The analysis would be similar, although P could then reduce the probability of control given to a bad type (from $(b-1)/(B-1)$ to $(b-1)/B$).

The optimal contract thus satisfies

$$p_g = 1, x_b = \frac{b-1}{B-1},$$

and the corresponding expected transfer p_b is

$$p_b = x_b B + b - p_g = \frac{b-1}{B-1} = x_b,$$

which can be implemented by asking for a price 1 whenever control is allocated to A.

This separating equilibrium yields 0 to a good A and $x_b(B-1) = b-1$ to a bad A, while P obtains a payoff of:

$$(1-\mu)\,2 - \mu\frac{b-1}{B-1}(l-1)$$

which is positive provided:

$$\mu \leq \mu_c^m = \frac{1}{1 + \frac{b-1}{B-1}\frac{l-1}{2}}.$$

■

In other words, when $b > 1$ the introduction of monetary transfers does not eliminate the role of (contractible) control allocation; in particular, it can be checked that, whenever the relationship is profitable ($\mu < \mu_c^m$), P prefers this separating contract to staying uninformed and keeping control over stage 1. However, the optimal separating contract still allocates control with positive probability to a bad A and never to a good A; thus, while monetary transfers allow a reduction in the probability of control given to a bad A required for truthtelling, the optimal contract still suffers from the same credibility problem as before. This again leads us to consider the case where the control allocation cannot be committed upon ex ante (at the contracting stage), even though control over stage 1 can be transferred ex post (given the messages possibly sent at the beginning of stage 1) to A.

5.2 Transferable control case

Allowing for monetary transfers in the transferable control case yields the following description of the strategy set and the timing of moves. In the contracting phase, P offers A a contract which stipulates messages and (possibly message-contingent) monetary transfers to be sent by A at the beginning of each stage. A then either refuses the contract, in which case both parties get zero, or accepts it, in which case the game proceeds as follows.

- In stage 1, A sends a message and makes a monetary transfer according to the contract; P then decides *with full discretion* whether or not to transfer control to the agent; finally, whoever ends up in control, decides upon the stage 1 action.
- In stage 2, A may send further messages with possibly new transfers according to the contract; finally, P decides whether or not to implement the project.

When $b > 1$, a bad type is more eager than a good one to get control over stage 1 and/or convince P to implement a project; however, P would never transfer control or implement the project if she learns that A is bad. It is therefore impossible to have A's revealed through type-contingent messages or payments:

Lemma 5 *Assume $b > 1$. When control over stage 1 is transferable but not contractible, then payments and control allocation cannot depend on the agent's type;.there is thus no loss of generality in not asking the agent to send messages.*

Proof. It suffices to show that A's payments and P's decisions cannot be message or type-contingent. We first show that there is no scope for revelation of A's type at stage 2. If separation were to occur at that stage, a bad A would not pay anything since an informed P would then stop the project. A good A could try to credibly report his type by making a specific payment p, in order to induce P to implement the project. However, a good A would be willing to pay only $p \leq 1$, and a bad A would then also report a good type since $p < b$.

We now show that there is no scope for messages at stage 1 either. In any separating equilibrium, P would keep control and stop the project when a bad A reveals his type. A bad A would thud not pay anything. A good A could try to reveal his type by making a specific payment p, in order to influence P's decision over the allocation of stage 1 control. But again a good A would be willing to pay p only if $p \leq 1$, in which case a bad A would pay p to obtain control and choose action N, even if this leads P to stop the project. ■

Given this lemma, at the contracting stage P must simply stipulate a price (which must be the same for both types) and choose between keeping control, in which case she remains uninformed about A's type by the end of stage 1, or transferring stage-1 control to A. The same analysis as in Section 4 then leads to:

Proposition 6 *Assume $b > 1$. When control over stage 1 is transferable but not contractible, the relationship is profitable as long as $\mu < \mu_t = 1/(1+l)$, in which case P's optimal contract is such that: (i) no messages are sent before stage 1 (and a fortiori after that stage); (ii) A pays 1 and P transfers control over stage 1 to A; (iii) in stage 1, a good A chooses C whereas a bad A chooses N, and in stage 2 P implements the project if and only if C has been chosen. In this optimal contract, A gets the same rents but P gets less than when control is contractible. If $\mu < \mu_t$, it is instead optimal for P to keep control and stop the project.*

Proof. The previous lemma establishes that P cannot learn A's type simply through message-contingent payments and the reasoning is then similar to that of section 4. If $\mu < \mu_t$, P prefers learning A's type by transferring control over stage 1. A good A then obtains 1 (by choosing C and inducing P to implement the project) and a bad A obtains $B > 1$ (it is a dominant strategy to choose N, even though it then leads P to stop the project). The maximal price P can ask is thus $p = \min(1, B) = 1$; a good A then gets no rent while a bad A gets $b - 1$, and P's expected payoff is

$$1 + (1 - \mu) - \mu l = (1 - \mu)\, 2 - \mu\, (l - 1)\,,$$

which is lower than when control is contractible.

If P opted instead for keeping control with probability 1, she would remain uniformed about A's type and implement the project only if $\mu < \mu^*$; in that case, she could require at most a price $p = 1$, which is the maximal price a good A is willing to pay. She would thus get $(1 - \mu)\, 2 - \mu(L - 1)$, which is lower than what she can get with the above revelation mechanism. And if $\mu > \mu^*$, P would stop the project and thus earn zero (since she could not ask for any positive payment in that case). Therefore, P prefers the above mechanism to keeping control with probability 1

If $\mu > \mu_t$, P prefers keeping control and stopping the project; anticipating this, A does not pay anything. ■

Introducing monetary transfers thus allows P to extract some of the agent's rents (when the relationship is profitable) but does not otherwise affect the optimal contract. Transferring control remains an effective instrument for trust-building – and the only way for P to learn A's type in stage 1.[14]

6 A foundation for the notion of transferable control

The previous section has shown the importance of the assumption of noncontractible-but-transferable control for the optimality of relying on control transfers as reputation devices. In this section, we investigate a precise justification for the assumption that control cannot be allocated contractually but can be transferred irreversibly nonetheless, provided it is in the interest of the individual who transfers it. In general, one can think of several justifications. For example, assume that undertaking the stage 1 action requires the acquisition of *skills and knowledge*, and party P is initially in charge of deciding who receives the required training, P or A; if the needed skills and knowledge are not easily verifiable, outside parties may be unable to assess who has received the required training, and P is in a position to decide who receives this training without being able to commit himself ex ante on that subject. More generally, the party in charge of a project may credibly transfer control to another party over specific tasks by dedicating her own *time* to other activities (at least in the short-run) and concentrate that other party's *time and attention* on that particular task (by reducing that party's involvement in other, nonverifiable tasks, by giving that party better instruments to operate that task, etc.).

Alternatively, we can connect the notion of transferable control and the con-

[14] When control is contractible, P can however do better using the above-described revelation mechanism, which minimizes the probablity of giving control to a bad A. It can be checked that, whenever $\mu < \mu_c^m$, the optimal contract in the contractible case is indeed the revelation mechanism described in the previous section; in particular, as already mentioned P gets more with the revelation mechanism than by staying uninformed and keeping control over stage 1; P also prefers the revelation mechanism to transferring control with probability 1 to A whatever his type. When $\mu > \mu_c^m$, the relationship is not profitable and P is better off keeping control over stage 1 and stopping the project in stage 2.

cepts of formal and real authority introduced by Aghion and Tirole (1997), hereafter AT. They stress the fact that the exercise of authority requires more than the *formal right* to make decisions: it may often also require *relevant information* in order to take appropriate decisions. AT investigate this issue in the context of an investment problem where one project has to be chosen among n ex-ante indistinguishable projects. Contracting first takes place over formal authority, then effort is exerted by the parties in order to acquire information about the payoffs of the various investment projects, and finally the investment project can be chosen. Since the parties have partially congruent payoffs, it may be in the interest of the party endowed with formal authority, when she is uninformed, to "grant" real authority to the other party, by following his recommendation.

We can use AT's basic setup to provide foundations for the notion of transferable control. More specifically, let us reinterpret stage 1 of our model in the light of the AT framework: label party P's favorite project as "action C", while A's favorite project is "action N". Knowing which actions are C and N is however not obvious, because they are among n possible actions that are ex ante indistinguishable. A's type in our model can be interpreted as the degree of congruence of A's payoff function with P's payoff function: a "good type" is congruent with P, while a "bad type" is not. The difference in information structure between AT and us is that AT assume that both parties are initially uninformed about project returns but know the degree of congruence of their payoffs. Here instead, P knows the various project returns, but does not know the degree of congruence between A's interest and her own. This matters because P has to rely on A in stage 2 if she does not stop the project. The question then is whether P can learn about A's payoff structure by giving her authority in stage 1.

In this setup, the notion of transferable control appears very naturally. Remember that what we require is that control cannot be allocated contractually but can be transferred irreversibly nonetheless, provided it is in the interest of the individual initially in control to do so. Assume that P allocates formal stage-1 decision rights to A at the beginning of the project. If informed about the project returns, A would then take the stage-1 decision that he prefers. However, if A is *initially uninformed* about the project returns, to make a stage-1 decision A will need information from P (assume as in AT that, except for C and N, all the other projects are so bad that choosing at random is worse than doing nothing at all). It is natural to assume that, in many instances, P *cannot commit ex ante to transmit the appropriate information.* In particular, in stage 1, P could decide to only tell A what action C is, and not action N. On the other hand, once P has transmitted the information, control has been *irreversibly* transferred: once A knows about action N, there is nothing to prevent him from choosing it, given that he has received formal stage-1 control rights initially.

Transferring "real control" in stage 1 implies in fact that P informs A about the identity of both "action C" and "action N". This transfer of real authority never occurs in AT: there, when the party endowed with formal authority is uninformed and turns to the other party for advice, it receives information about a *single project* from that party,[15] and therefore has no choice but to follow the recommendation. In AT, an informed party never transfers real authority, because getting one's favorite project implemented is the only thing that counts at this stage. Here instead, although P is fully informed about project returns, she finds it optimal to transfer real authority to A in stage 1 in order to learn about his type, which is helpful for stage 2. *We therefore have true delegation here, motivated by the concern for future cooperation.*

With this reinterpretation of the stage-1 actions along the lines of the AT setup, we can now apply straightforwardly the results of the previous section: provided that the probability μ that A is not congruent enough with P is not too large, it can become optimal for P to give away (or sell, under monetary responsiveness) formal authority over the stage-1 action to A, and also to transmit information to A about which among the n actions are actions C and N. Through this transfer of control, P has the opportunity to learn about A's type, and to stop the project in stage 2 if and only if A has chosen action N in stage 1.

7 Application: the scope of delegation[16]

So far, we have restricted the stage-1 action set to only two actions, C and N. We now extend this set to include convex combinations of these two actions. Specifically, for any $\alpha \in [0, 1]$, action N_α generates payoffs $-\alpha l$ for the principal, 0 for a good agent and αB for a bad agent. We shall now allow the principal

[15] Or zero, if the other party in also uninformed.

[16] For a potential application of our notion of transferable control to the analysis of delegation within large organizations, see Hart-Holmstrom (2002).

to grant partial control to the agent, over any subset of actions N_α. In the AT interpretation, this amounts to give the agent the relevant for these actions. An alternative interpretation is that stage 1 consists of several projects, each of which involves two actions N and C, with payoffs as described above. The principal then decides which fraction of projects to delegate to the agent.

This convexification of the action set allows the principal to replicate the contractible control optimum with only transferable control. Namely, it is optimal for the principal to transfer control over the actions N_α for $\alpha \in [0, \bar{\alpha}]$, where $\bar{\alpha} = \frac{b}{B}$ represents the minimal amount of delegation required for the bad agent to reveal his type. The extent of delegation thus increases with B and decreases with b: since the goal is to induce the bad agent to signal his type by choosing action N, it is easier, the higher the short-term gain of doing so, i.e. B, and the smaller the long-term loss of doing so, i.e. b.

This can be naturally related to the mobility of the environment:

- Suppose for example that, with exogenous probability ρ, the principal-agent relationship disappears after stage 1, e.g., due to the technological obsolescence of the firm or of the agent's firm-specific skills. Then b must be replaced with $(1-\rho)\, b$ and the optimal amount of delegation becomes $\bar{\alpha}(\rho) = (1-\rho)\frac{b}{B}$. Thus, delegation decreases with the rate of obsolescence.

- Alternatively, one could assume that the benefit αB that the bad agent obtains from action N_α now only arises if the agent receives an outside job offer. This would for example be the case if action N corresponds to an investment in the agent's general human capital or, more generally, in his market value, at the expense of the firm. Call this probability ρ'. Now, the optimal amount of delegation is defined by $\alpha = \frac{b}{B}/\rho'$. Thus, delegation decreases with labor market mobility as measured by ρ'.

References

Aghion,. P. and P. Bolton (1992), "An Incomplete Contracts Approach to Financial Contracting," *Review of Economic Studies* 59: 473-494.

Aghion, P., M. Dewatripont and P. Rey (1994), "Renegotiation Design with Unverifiable Information," *Econometrica* 62: 257-282.

Aghion, P., M. Dewatripont and P. Rey (2002), "On Partial Contracting," *European Economic Review*, forthcoming.

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

Baker, G., R. Gibbons and K.J. Murphy (1997), "Relational Contracts and the Theory of the Firm," mimeo, MIT.

Banerjee, A. and E. Duflo (1999), "Reputation Effects and the Limits of Contracting: A Study of the Indian Software Industry," mimeo, MIT.

Boot, A., S. Greenbaum and A. Thakor (1993), "Reputation and Discretion in Financial Contracting," *American Economic Review* 83: 206-212.

Dessein, W. (2000), "Control Allocations in Joint Undertakings: Balancing Formal and Real Authority", mimeo.

Dessein, W. (2002), "Authority and Communication within Organizations," *Review of Economic Studies*, forthcoming.

Dewatripont, M. (2001), "Authority," Walras-Bowley Lecture presented at the North American Summer Meeting of the Econometric Society, College Park, Maryland.

Dewatripont, M. and J. Tirole (1994), "A Theory of Debt and Equity: Diversity of Security and Manager-Shareholder Congruence," *Quarterly Journal of Economics* 109: 1027-1054.

Grossman, S. and O. Hart (1986), "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy* 94: 691-719.

Halonen, M. (1997), "A Theory of Joint Ownership," mimeo, University of Bristol.

Hart, O. (1995), *Firms, Contracts and Financial Structure*, Oxford: Oxford University Press.

Hart, O. and J. Moore (1988), "Incomplete Contracts and Renegotiation," *Econometrica* 56: 755-786.

Hart, O. and J. Moore (1990), "Property Rights and the Nature of the Firm," *Journal of Political Economy* 98: 1119-1158.

Hart, O. and J. Moore (1999a), "Foundations of Incomplete Contracts," *Review of Economic Studies* 66: 115-138.

Hart, O. and J. Moore (1999b), "On the Design of Hierarchies: Coordination versus Specialization," mimeo, Harvard and LSE.

Hart, O. and B. Holmström (2002), ""Vision and Firm Scope", mimeo, Harvard and MIT.

Holmström, B. (1979), "Moral Hazard and Observability," *Bell Journal of Economics* 10: 74-91.

Holmström, B. (1982), "Moral Hazard in Teams," *Bell Journal of Economics* 13: 324-340.

Kreps, D., P. Milgrom, J. Roberts and R. Wilson (1982), "Reputation and Imperfect Information," *Journal of Economic Theory* 27: 253-279.

Legros, P. and S. Matthews (1993), "Efficient and Nearly Efficient Partnerships," *Review of Economic Studies* 60: 599-611.

Legros, P. and A. Newman (1999), "Competing for Ownership," mimeo, ECARES.

Maskin, E. (1999), "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies* 66: 23-36.

Maskin, E. and J. Moore (1999), "Implementation and Renegotiation," *Review of Economic Studies* 66: 39-56.

Maskin, E. and J. Tirole (1999a), "Unforeseen contingencies and Incomplete Contracts," *Review of Economic Studies* 66: 83-114.

Maskin, E. and J. Tirole (1999b), "Two Remarks on the Property-Rights Literature," *Review of Economic Studies* 66: 139–149.

Mirrlees, J. (1999), "The Theory of Moral Hazard and Unobservable Behavior, Part I," *Review of Economic Studies* 66: 3-22.

Moore, J. and R. Repullo (1988), "Subgame Perfect Implementation," *Econometrica* 56: 1191-1220.

Noldeke, G. and K. Schmidt (1995), "Option Contracts and Renegotiation: A Solution to the Hold-Up Problem," *Rand Journal of Economics* 26: 163-179.

Segal, I. (1999), "Complexity and Renegotiation: A Foundation for Incomplete Contracts," *Review of Economic Studies* 66: 57-82.

Tirole, J. (1999), "Incomplete Contracts: Where Do We Stand?," *Econometrica* 67: 741-781.

Watson, J. (1999), "Starting Small and Renegotiation," *Journal of Economic Theory* 85: 52-90.