Transferable control

Philippe Aghion, Mathias Dewatripont and Patrick Rey[1]

July 2003

[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, participants in the International Francqui Conference (Brussels, November 1999) on the Economics of Contracting, various seminar audiences, and especially three referees for very useful comments and suggestions.

Abstract

In this paper, we introduce the notion of *transferable control*, defined as a situation where one party (the principal, say) can transfer control to another party (the agent) but cannot commit herself to do so. One theoretical foundation for this notion of transferable control builds on the distinction between formal and real authority introduced by Aghion and Tirole (1997), in which the actual exercise of authority may require noncontractible information, absent which formal control rights are vacuous. We then use this notion to study the extent to which control transfers may allow an agent to reveal information regarding his ability or willingness to cooperate with the principal in the future. We show that the distinction between contractible and transferable control can drastically influence how learning takes place: with contractible control, information about the agent can often be acquired through revelation mechanisms that involve communication and message-contingent control allocations; in contrast, when control is transferable but not contractible, it can be optimal to transfer control unconditionally and learn instead from the way in which the agent exercises control.

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. However, this assumption is not always warranted. For example:

- The President of a country like France cannot contractually commit not to change (that is, withdraw control from) his/her Prime Minister: Even if reputation considerations make it credible that he/she will not fire the new Prime Minister right after having appointed him/her, this obviously changes subsequently depending on the Prime Minister's performance on the job..

- More generally, the promotion/demotion of a subordinate in an organization (that is, the withdrawing/granting of control for a given set of tasks) rarely involves prior contractual commitment, but instead relies on (often soft) information about job performance.

- As for the provision of new credit lines by banks or credit card companies, there is also typically no contractual commitment for these lines to be maintained, as banks want to protect themselves against possible abuses by customers; at the same time, credit lines will not be withdrawn right after having been granted, and control over the future course of action lies with the customer until the credit line is removed.

All these examples have the following two features in common: (i) control is clearly not fully contractible ex ante, but can be transferred somewhat irreversibly, at least in the short run *when the party doing so finds it in his/her interest*; (ii) putting a party "in control" allows the other party to test and learn

[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), Dessein (2000, 2002), Hart and Moore (1999b) and Hart-Holmström (2002).

more about the agent's ability or loyalty to the organization. In this paper, we therefore explore the implications of *transferable control*, defined as a situation where one party (the principal, say) can transfer control to another party (the agent) but cannot commit herself to do so. One theoretical foundation for this notion of transferable control builds on the distinction between formal and real authority introduced by Aghion and Tirole (1997), in which the actual exercise of authority may require certain critical information, absent which formal control rights are vacuous;[4] and while formal control rights could be contracted upon, information transfers may not (if for instance the informed party claims that she has no useful information, or provides useless information, when she does not consider the information transfer to be in her interest).

In this paper, we use this notion to study the extent to which control transfers may allow an agent to reveal information regarding his willingness to cooperate with the principal. We show that the distinction between contractible and transferable control can drastically influence how learning takes place: with contractible control, information about the agent can often be acquired through revelation mechanisms that involve communication and message-contingent control allocations; in contrast, when control is transferable but not contractible, it can be optimal to transfer control unconditionally and learn instead from the way in which the agent exercises this control.

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: contracts can then fully determine the entire interaction between the parties; this case encompasses the *implementation literature* à la Maskin (1999) or Moore-Repullo (1988), where one can contract on entire game forms. In such a world, who "chooses" the action is irrelevant: the contracting parties can limit themselves to sending (possibly sequential) messages to a "planner" who then takes or dictates all relevant actions.[5]

[4] Alternatively, undertaking an action may require the acquisition of unverifiable *skills and knowledge*. The party initially in charge may decide to transfer the required skills (or provide the adequate training) without being in a position to contractually commit to do so.

[5] The implementation results of this literature have been generalized by Maskin-Tirole (1999) to the case where actions are noncontractible "ex ante", before the revelation of the state of

(ii) At the other extreme, a world with only noncontractible, pre-assigned actions: here, contracts hardly affect the structure of the game played by the parties; this case encompasses both *game-theoretic models* (such as Kreps et al. (1982), Sobel (1985) and Watson (1999)), which assume away any contracting, and *moral hazard models* (à la Mirrlees (1999), Holmström (1979, 1982), Legros-Matthews (1993)), where noncontractible actions are pre-assigned to one or the other party and can only be influenced indirectly, by contracting over related variables (e.g., output).

In-between, there is *partial contracting* (see Aghion et al. (2002)), where formal contracts do not determine the entire relation between the contracting parties but can influence the *underlying game* between them.[6] These are situations in which some actions are ex post nonverifiable and therefore not contractible ex ante, so that they cannot be delegated to (or dictated by) a social planner. Yet *control* over such actions can influence the dynamic interaction between the parties. Much of the existing literature on control rights and authority has concentrated on situations where such control is fully contractible.[7]

Our paper is first related to the above literature on control rights and authority, to which we add the possibility of credible control transfers. We thus focus on a case intermediate between contractible control and pre-assigned actions (or "moral hazard"), where control over particular actions is not contractible but still credibly transferable. Second, our model also relates to the game-theoretic literature on reputation; allowing for contracting before the actual game takes place however allows us to study under which conditions – e.g., contractible versus transferable control – information is transmitted by action choices rather than through revelation mechanisms. Third, since we focus on control allocation as a way to induce cooperation, our analysis also relates to the literature on "formal" versus "informal" contracting, and in particular to Baker et al. (2002) and Halonen (1997). In a repeated model of ownership allocation à la Grossman-Hart (1986), where all parties have complete information, Baker et al. (2002) show

nature, but become contractible "ex post", after the revelation of the state of nature.

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

[7] See e.g. Aghion-Bolton (1992), Dewatripont-Tirole (1994), Aghion-Tirole (1997), Legros-Newman (1999) and Hart-Moore (1999b). See also Dewatripont (2001).

that vertical integration may help the parties to hold 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 will then find it particularly costly for their reputation 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, and they restrict attention to simple contracts, even though revelation mechanisms would be more effective (for example, in Halonen's framework, where trade is supposed to be ex post verifiable, relatively simple contracts could be quite powerful).

The paper is organized as follows. Section 2 describes the framework. Section 3 focuses on contractible control and shows that in a broad range of situations the optimal contract is a revelation mechanism promising control to the agent when he announces a non-cooperative type. This kind of mechanism ceases to be credible when control is not contractible but only transferable. In this case, as shown in Section 4, the power of revelation mechanism is greatly reduced and unconditional control transfers emerge as an optimal learning device. While Sections 3 and 4 went for simplicity in not allowing for monetary responsiveness, Section 5 shows that our results are robust to its introduction. Finally, Section 6 connects our analysis to the Aghion-Tirole (1997) concepts of formal and real authority, briefly discusses its implications for the study of delegation, and suggests some obvious extensions for future research.

2 Framework

This section outlines an incomplete information framework where control allocation serves as a natural instrument to enhance trust and cooperation. Specifically, we consider a relationship between a principal (she) and an agent (he), hereafter P and A, meant to carry out a project: P has overall control over the project

but needs A for implementing it. More precisely, the project involves two stages, design and implementation. In the design stage, the party in charge 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 A may favor – for example, it may enhance A's human capital or, more generally, A's market value. In the implementation stage, P decides whether to implement the project (I), or to stop it (S). 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 P and A can contract over who is in charge of design and the case where P can simply *transfer* control to A, without contractually committing herself to do so. The design decision itself (C or N) is observed by both parties but not contractible. We normalize to zero the parties' payoffs from C; adopting instead the non-cooperative action N does not affect a good A but entails a loss $(-l)$ for P and a benefit B for a bad A. P's and A's payoffs from N are thus respectively:[8]

$$(-l, 0) \text{ when } A \text{ is good,}$$
$$(-l, B) \text{ when } A \text{ is bad.}$$

- in stage 2, P freely decides whether to implement or stop the project (actions I and S respectively): this decision is not contractible and cannot be delegated to A. Stopping the project yields zero payoffs for both parties. Implementing the project brings instead an additional gain G to P and g to A when A is good, while it brings a loss $(-L)$ to P and a gain b to A if A is bad. P's and A's payoffs from I are thus respectively:

$$(G, g) \text{ when } A \text{ is good,}$$
$$(-L, b) \text{ when } A \text{ is bad.}$$

[8] At the end of Sections 3 and in Section 5, we discuss the robustness of our results with respect to changes in the payoff matrix.

We assume away discounting between the two stages; P's and A's overall payoffs are thus simply the sum of the first- and second-stage payoffs and can be summarized as follows:

- when A has a good type:

Action	I	S
C	G, g	$0, 0$
N	$G - l, g$	$-l, 0$

- when A has a bad type:

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

Figure 1

We shall restrict attention to the case where:

$$\begin{aligned} B &> b > 0, \\ L &> l > 0. \end{aligned}$$

Thus a good A is willing to cooperate in stage 1 and gains g 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 ($B > b$). 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).

This payoff structure is that of a typical signalling game, where preference heterogeneity between the two types of A can allow for separation. It is however particular in one important respect: there is congruence in both periods between the preferences of P and the good type of A (in particular, both prefer C followed

by implementation to N followed by stopping the project).[9] At the end of Section 3, we discuss the role of this congruence.

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 (whatever action has been chosen in stage 1 since utilities are all separable over time) whenever:

$$(1-\mu)G + \mu(-L) < 0;$$

(the left hand side is the expected stage-2 payoff of the principal if she implements the project, the right hand side is her stage-2 payoff if she stops the project); this can be reexpressed as:

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

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

We start our analysis by assuming that payoffs are private benefits and that the parties are not responsive to monetary incentives. Therefore, contracts 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. When parties are instead responsive to monetary incentives (in Section 5 below), optimal contracts also include message-contingent transfer payments.

[9] Fixing the implementation decision, a good A is here indifferent between cooperating or not on design. The analysis applies unchanged when a good A gains ε from adopting N, where ε is small but either negative (a good A strictly prefers to cooperate) or positive (a good A 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., $g > \varepsilon$).

3 Contractible control: The power of revelation mechanisms

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 absence of monetary responsiveness).

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 (possibly randomly) 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 however 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.

For simplicity, we concentrate below on contracts where pure strategies are played (and we explain in footnotes how the results are robust to the possibility of mixed strategies). 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 over the stage 1 action, and implement the project in stage 2 if and only if $\mu < \mu^*$. P's payoff is then $\max\{(1-\mu)G - \mu L, 0\}$. 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.*[10]

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 g 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

$$(1-\mu)\,G - \mu y l,$$

[10] This result remains valid when equilibrium mixed strategies are considered, which can be seen as follows: (i) first, if A obtains control in stage 1, he will play a pure strategy for sure, and so will P in stage 2; (ii) if P is in control in stage 1, she will not mix in stage 2 (between I and S) if her posterior probability assessment that A is bad is not equal to μ^*. But if P does not mix in stage 2, she prefers the bad A also not to mix (between saying he is bad or good), since the loss from I (L) outweighs the loss from stage 1 control (as for the good A, he never wants to mix, since he is indifferent between C and N, and only cares about the stage 2 action); (iii) finally, an outcome where P is in control in stage 1 and mixes in stage 2 is also unattractive for P, because it means she obtains a zero payoff upon having stage 1 control (and a negative one upon not having stage 1 control).

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

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

and is thus positive whenever

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

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

$$(1-\mu)\, G - \mu \frac{b}{B} l > (1-\mu)\, G - \mu L.$$

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

Remark: We stressed above that the payoff structure was particular in having P's preferences congruent with that of a good A. Indeed, the above revelation mechanism no longer works when the principal's preferences become sufficiently non-congruent with those of *both* types of agents. Suppose for example[11] that, at stage 1, there are three possible actions to be taken: C, N and N'. The payoffs from C and N are the same as before, and N' brings zero to P and a bad A; but a good A now gains $B_g > 0$ from N'. Therefore, P prefers C to N and N', and a bad A prefers N to C and N', but a good A now prefers N' to C and N.

[11] We thank a referee for suggesting this example.

Let as before x and y denote the probabilities that A obtains control when announcing a good and a bad type. To induce A to tell the truth, we must have:

- for a good type:
$$xB_g + g \geq yB_g + yg; \tag{1}$$
- for a bad type:
$$yB \geq xB + (1-x)b. \tag{2}$$

Indeed, if a good A report his type truthfully, in stage 1 he obtains control with probability x and can then choose his preferred action N', while in stage 2 P implements the project; if he reports instead a bad type, with probability y he gets control and reveals his true type by choosing N', inducing P to implement the project,[12] otherwise P chooses C and then stops the project. Similarly, if a bad A reports his type truthfully he obtains control and chooses N with probability y, while P stops the project to avoid losses in stage 2; if he pretends instead to be a good type, with probability x he acquires control and then reveals his type by choosing N, otherwise the principal keeps control and implements the project.

The two incentive conditions can be rewritten as

$$(1+\alpha)(1-y) \geq 1-x \geq \frac{1}{1-\beta}(1-y),$$

where $\alpha = g/B_g$ and $\beta = b/B$. Therefore, if

$$(1+\alpha) < \frac{1}{1-\beta},$$

both conditions can be simultaneously satisfied only if $x = y = 1$, which amounts to simply transferring control to A and learning from his choice of action.

4 Transferable control and learning by delegation

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

[12] Alternatively, P could stick to the belief that A has a bad type; however, such a belief would contradict for example Cho and Kreps' intuitive criterion, since only a good A gains from choosing N', even if doing so is the only way to induce P to implement the project.

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. The set of strategies and the timing of events are then 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, 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 his good type and ensures that P implements the project.[13] In comparison with the case where 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

[13] There may also exist pooling equilibria where both types of agents choose action N . These equilibria are however dominated by those in which P keeps control over the stage 1 action.

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 G if A turns out to be a good type. Thus, learning A's type when $\mu > \mu^*$ generates an additional expected gain equal to $(1-\mu)\,G$ to the principal; for such values of μ, P will thus prefer to grant control to A in stage 1 if:
$$(1-\mu)\,G > \mu l,$$
or equivalently:
$$\mu < \mu_t \equiv \frac{G}{G+l}.$$
This establishes our main result:

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.*[14]

[14] Once again, mixed strategies will not alter the result: whatever the equilibrium messages

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's 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 reveal his willingness to cooperate, 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 to induce 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.[15]

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:

$$(1-\mu)\,G - \mu l.$$

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:

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

Therefore:

- 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

sent by A, the probability that P transfers control cannot be increasing in her probability assessment that A is bad. On the other hand, it is also impossible that it be strictly decreasing in equilibrium, because this would violate the bad A's incentive constraint. Therefore, messages can be ignored and we are left with unconditional control transfers.

[15] On the other hand, P must be able to commit to let A exert control over a number of actions and/or for some time, so as to allow a bad A to gain sufficiently from the exercise of this control. If for example P had the right to "withdraw" control from A at any moment, then P would indeed be tempted to overrule A's choice of N. Anticipating this, a bad A would then "choose" C to preserve his benefit from implementation and the principal would thus never learn the agent's "exercise" of control. More generally, P must be in a position to transfer control irreversibly for at least some time, for such control to be useful as a learning device.

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.

- When control is transferable but not contractible, the relationship is profitable only when
$$\mu < \mu_t = \frac{G}{G+l},$$
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 (defined as in the previous section), and p is a monetary transfer from A to P. Since we now have transferable utilities, what matters for efficiency is the sum of the parties' payoffs; in keeping with the analysis of the previous sections, we assume in this section
$$L > l > B > b,$$
so that it is efficient to stop the project when the agent is bad.

Allowing for monetary responsiveness means that contracts can require transfers contingent on messages, which in turn can provide additional ways to acquire the information over A's type. In particular, if $b < g$, P can simply keep control and obtain full revelation at stage 2 by having the good A pay g against project implementation.[16] The possibility of monetary transfers thus eliminates the role for allocating or transferring control to A in that case. When $b > g$, however, the previous insights remain valid, as we show below.

[16] If $b < g$, 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 $p = g$; (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.

5.1 Contractible control

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 $b > g$. Then, the optimal separating contract for P consists in selling control at price g with probability $(b-g)/(B-g)$ when A reports a bad type, while keeping control and requiring a payment of g when A reports a good type.*[17] *In this mechanism, a good A obtains no rents while a bad A obtains $b-g$ and P obtains a positive expected payoffs as long as*

$$\mu < \mu_c^m \equiv \frac{1}{1 + \frac{b-g}{B-g}\frac{l-g}{G+g}}.$$

Proof. See Appendix. ■

In other words, when $b > g$ the introduction of monetary transfers does not eliminate the role for (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 for a reduction in the probability of control given to a bad A, the optimal contract still suffers from the same credibility problem as in Section 3.

Remark: As shown in Aghion et al. (2003), allowing for monetary transfers tends to increase the power of revelation mechanisms: Charging for control allocation reduces a good A's net gain from acquiring control, and therefore reduces

[17] Once again, this result is robust to the consideration of mixed strategies. Essentially the same arguments apply as in footnote 10, since the mechanism is very similar - although with a different probability of giving control to A - except that there is a price g to be paid by A whenever he actually receives stage 1 control: (i) first, if A obtains control in stage 1, he will play a pure strategy for sure, and so will P in stage 2; (ii) if P is in control in stage 1, she will not mix in stage 2 (between I and S) if her posterior probability assessment that A is bad is not equal to μ^*. But if P does not mix in stage 2, she prefers the bad (resp. good) A also not to mix (between saying he is bad or good), because I (resp. S) is inefficient; (iii) finally, an outcome where P is in control in stage 1 and mixes in stage 2 is also unattractive for P, because, on the one hand, it means she does not benefit from the gain of implementation of the project and, on the other hand, by lowering the probability of implementation, she will moreover obtain a lower expected payment from the good A (whose participation constraint is binding) and in turn from the bad A (whose incentive constraint is binding) at the initial message stage.

the degree of non-congruence between a good A and P. Yet, we show that these revelation mechanisms only work if the lack of congruence between the good A's preferences and P's preferences is not too severe.

5.2 Transferable control

When $b > g$, 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 learned that A was bad. It is therefore impossible to have A's type revealed through type-contingent messages or payments. In fact we can show:

Lemma 5 *Assume $b > g$. 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. See Appendix. ■

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

Proposition 6 *Assume $b > g$. When control over stage one is transferable but not contractible, the relationship is profitable as long as $\mu < \mu_t = G/(G + l)$, in which case P's optimal contract is such that: (i) no messages are sent before stage one (and a fortiori after that stage); (ii) A pays g and P transfers control over stage 1 to A; (iii) in stage one, a good A chooses C whereas a bad A chooses N, and in stage two P implements the project if and only if C has been chosen in stage one. In comparison with the contractible control optimum, the good A still gets no rents, the bad A gets higher rents and P gets lower rents. Finally, if $\mu \geq \mu_t$, it is optimal for P to keep control and stop the project.*[18]

[18] Just as without monetary responsiveness, mixed strategies will not alter the result: whatever the equilibrium messages sent by A, the probability that P transfers control cannot be increasing in her probability assessment that A is bad. On the other hand, it is also impossible that it be strictly decreasing in equilibrium, because this would violate the bad A's incentive

Proof. See Appendix. ■

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 the only way for P to learn A's type in stage 1.[19]

6 Discussion and conclusions

This paper provides contract theoretic foundations for games in which a principal needs to transfer control rights to her agent in order to "test" her ability or propensity to cooperate in the future. Specifically, we have shown that when control is transferable, as opposed to being contractible, simple control transfers emerge as optimal learning devices. Thus, moving from contractible to transferable control can significantly reduce the power of revelation mechanisms: the principal may no longer rely on communication and message-based control transfers, but simply put the agent in charge and learn from the agent's exercise of control.

Taking a contractual perspective moreover generates additional insights for the theory of organizations. In this section, we discuss two applications of our framework to, respectively, the transfer of "real" authority and the scope of delegation.

6.1 Transferring real authority

Aghion and Tirole (1997) (hereafter AT) stress that the exercise of authority requires more than the *formal right* to make decisions: it often also requires *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

constraint. Moreover, the condition $b > g$ also rules out any separation between types, even a probabilistic one, using monetary payments. Therefore, we are left with unconditional control transfers and monetary payments.

[19] However, when control is contractible, P does better using the revelation mechanism described in subsection 5.1, which minimizes the probability of giving control to a bad A.

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.

In order to connect AT's basic setup to our analysis, 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 him (real) authority in stage 1.

Let us first assume that the formal stage-1 decision rights necessarily belong to A (suppose for example, as in the classical moral hazard literature, that P is either too busy or unable to actually take the action in that stage). In this case, the relevant notion of "control" is, who has the information needed to take decision, that is, who has real authority. And to the extent that the relevant information is not "verifiable", this control is not contractible but it *is* transferable: since A has formal authority, providing him with the relevant information indeed amounts to an irreversible transfer of real authority.

To be specific, 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. Since A is initially uninformed about project returns, he then needs information from P to make any stage-1 decision. It is natural to assume that, in many instances, P *cannot commit ex ante to transmit the appropriate information.*[20] On the other hand, if P informs A about the identity of both "action C" and "action N",

[20] In particular, if contractually instructed to identify x actions for A, P could decide to only tell A about action C and "bad" actions (possibly identifying several times the same action), but not action N.

control is *irreversibly* transferred to A: 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.

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.

Let us now assume that formal stage-1 decision rights can be contractually allocated to either P or A. Two cases might then be considered:

- if P can provide A with the relevant information about C and N *before* allocating formal stage-1 decision rights, we are back in the case of contractible control: indeed, by informing A beforehand, P can credibly grant real authority through the (possibly message-contingent) allocation of formal stage-1 decision rights. As in section 3, P may then find it optimal to first provide A with the relevant information and, second, offer formal control to A with positive probability when reporting a bad type.

- if instead the relevant information about C and N cannot be understood by A until *after* he has already been allocated formal authority over stage 1, then the logic of transferable control applies again. Contractually allocating formal authority to A is indeed no longer a credible way to grant him real authority. However, once formal authority has been allocated to A, P can also grant him real authority by providing the relevant information. The analysis is then essentially the same as in the beginning of this section, where A was always endowed with formal stage-1 control.

6.2 The scope of delegation

In the previous subsection, we identified circumstances under which P informs A about both actions C and N, in order to let him reveal his willingness to cooperate. In contrast, in AT, no such dynamic consideration exists and implementing

one's favorite project is the only thing that matters. As a result, transfers of real authority never occur: when the party endowed with formal authority turns to the other party for advice, it receives information about *at most one project*, and therefore has no choice but to follow the recommendation. In our setup, P finds it instead 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.

We can build on this insight to analyze the optimal scope or *size* of delegation. To this end, we extend the action set to include convex combinations of actions C and N. 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. P can now grant partial control to A over any subset of actions N_α. In the AT interpretation, this amounts to giving A the relevant information about 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. P then decides which fraction of projects to delegate to A.

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

This extension leads for example to intuitive comparative statics results:

- 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 $\bar{\alpha}(\rho) = (1-\rho)\, b/B$. 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 $\alpha = b/(\rho' B)$. Thus, delegation decreases with labor market mobility as measured by ρ'.

6.3 Further possible extensions

We have focused on a simple model with two periods, two types of agent and two actions. Yet, we have allowed for a full-blown mechanism design approach, in which the contracting set becomes particularly rich in the case with monetary responsiveness. The main insights should generalize to the case of multiple types and actions, but investigating optimal learning through control allocation in a dynamic setting is an interesting avenue for further research.

The above discussion illustrates the potential for the notion of transferable control to enrich the analysis of delegation in organizations in a dynamic context. In particular, we have been able to rationalize transfers of real authority as a learning device. This could be a useful building block for further investigating the interaction between information flows, the design of hierarchies and trust-building in organizations.

7 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 (2002), "On Partial Contracting," *European Economic Review* 46(4-5):745-753.

Aghion, P., M. Dewatripont and P. Rey (2003), "Transferable Control," mimeo.

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 (2002), "Relational Contracts and the Theory of the Firm," *Quarterly Journal of Economics* 117(1): 39-84

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* 69: 811-838.

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 B. Holmstrom (2002), "Vision and Firm Scope", mimeo Harvard/MIT.

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 (1994), "A Theory of Debt Based on the Inalienability of Human Capital," *Quarterly Journal of Economics* 109: 841-879.

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.

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

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.

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

Sobel, J. (1985), "A Theory of Credibility," *Review of Economic Studies* 52: 557-573.

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.

8 Appendix

8.1 Proof of Proposition 4:

As in section 3, we can restrict attention to direct mechanisms where, at each stage, A sends one of 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

$$g - p > 0,$$

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 $b > g$.

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,[21] and P then implements the project.

The two participation constraints are thus:

$$g - p_g \geq 0$$

and:

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

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

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 the project. When A gets control, choosing N would again lead P to stop the 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:[22]

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

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

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

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 $B \geq b > g$, 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:

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

and:

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

[22] 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 $(b-g)/(B-g)$ to $(b-g)/B$).

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

$$p_g + (1-\mu)\,G - \mu\left[x_b\,(l-B) + b\right],$$

while the first two conditions become

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

The optimal contract thus satisfies

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

and the corresponding expected transfer p_b is

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

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

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

$$(1-\mu)\,(G+g) - \mu \frac{b-g}{B-g}(l-g)$$

which is positive provided:

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

This establishes the Proposition.

8.2 Proof of Lemma 5

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 $p \leq g$, 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 thus 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 $p \leq g$, 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. The lemma is proved.

8.3 Proof of Proposition 6

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 g (by choosing C and inducing P to implement the project) and a bad A obtains $B > g$ (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 $p = \min(g, B) = g$; a good A then gets no rent while a bad A gets $B - g$, and P's expected payoff is

$$g + (1 - \mu) G - \mu l = (1 - \mu) (G + g) - \mu (l - g),$$

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 still require at most a price $p = g$, which is the maximal price a good A is willing to pay. She would thus get $(1 - \mu) (G + g) - \mu(L - g)$, 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.

This completes the proof.