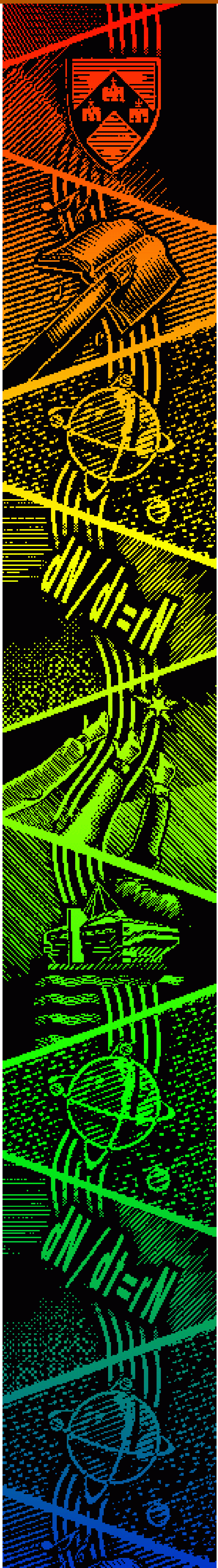

THE UNIVERSITY of York

Discussion Papers in Economics

No. 2004/10

Delegating Budgets when Agents Care About Autonomy

by

Michael Kuhn

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Delegating budgets when agents care about autonomy

Michael Kuhn[*]

University of York, Department of Economics and Related Studies and Centre for Health Economics[+]

May 2004

Abstract

We consider resource allocation within an organisation when agents have a preference for autonomy and show how delegation bears on moral hazard and adverse selection. Agents may care about autonomy for reasons of job-satisfaction, status or greater reputation of performance under autonomy. Separating allocations (overall budget and degree of delegation) are characterised depending on the preference for autonomy. As the latter is increasing, the degree of delegation assigned to productive and unproductive agents becomes more similar and may even be reversed when financial transfers are used. If agents' preference for monetary rewards is sufficiently weak, the principal will not employ financial transfers and pooling arises if the preference for autonomy is strong.

Keywords: adverse selection, capital budgeting, delegation, moral hazard, non-responsiveness, resource allocation.

JEL-classification: D 82, G31

[*] I am indebted to Luigi Siciliani and Peter Smith as well as to the audiences of EARIE 2002 and seminars in York and Bergen for helpful comments on earlier versions of this draft. The responsibility for remaining errors lies entirely with me.

[+] Heslington, York YO10 5DD. E-mail: mk19@york.ac.uk.

1 Introduction

When devising resource (or budget) allocations, principals regularly face a trade-off between delegating to an agent tasks, for which he has a comparative advantage, and retaining sufficient control of the agent's actions. This control is warranted to contain the potential problems of adverse selection – the agent over-stating the value of the project or his own productivity in order to attract additional budget – and moral hazard – the agent providing insufficient effort in using the budget, leading to under-achievement. In a perfect world, the principal could address these agency problems either by retaining full control or by delegating as fully as possible, but at the same time exposing the agent to the strongest possible incentives, in the extreme by selling the firm. In reality, the presence of monitoring costs on the one hand and wealth constraints or risk aversion on the other render these solutions impractical. Usually the principal will have to strike a balance between delegating and retaining control. Harris and Raviv (1998) explain delegation of budgeting decisions as a response to costly auditing. The cost of communication, the incompleteness of agency contracts and the 'value of flexibility' also tend to favour delegation (Melumad et al. 1997). Holmstrom and Roberts (1994) demonstrate that delegation - in the sense of transferring assets to the agent and/or of relaxing controls on the returns of assets accruing to the agent - paired with strong outcome related incentives, is an optimal response to a lower cost of outcome measurement or a lower risk. In contrast, more control is favoured if spill-overs between agents require central co-ordination or if agents hold lower bargaining power than the principal vis-à-vis third parties (Caillaud et al. 1996).

One common aspect of these models is that agents are concerned about the degree of control only to the extent to which it limits their scope to manipulate information or engage in slacking. In this regard, the models are firmly rooted in the neo-classical paradigm in which agents are motivated by extrinsic incentives, i.e. by performance-related rewards or punishments. A more recent line of literature acknowledges that agents may be motivated by non-financial-aspects of their jobs. Frey (1993, 1997) argues that agents may be driven by intrinsic motivation. This means that agents derive satisfaction from performing a task well and provide effort even in the absence of extrinsic incentives.[1] However, intrinsic motivation is not independent of the working environment. Specifically, it may be crowded out by extrinsic incentives such as performance standards and the associated punishments or rewards.[2] The latter tend to destroy the agents' self-evaluation of doing 'something decent' over and above what is expected or even enforced anyway. The control associated with extrinsic rewards also leads to a loss of self-determinedness that further undermines motivation.[3] The possible crowding-out of intrinsic motivation constrains the principal's scope in providing external performance incentives.[4]

[1] Other non-financial motivations include altruism, as often assumed in the modelling of physician behaviour (e.g. McGuire 2001: section 6.2), and status (e.g. Encinosa et al. 1997).

[2] Hence, agents not only care about outcomes but also about the process at which these outcomes are arrived at (Sen 1997). Frey and Benz (2002) argue that procedural utility derived from the mode of production (self-employed or not) matters and provide evidence supporting the view that job satisfaction decreases in the degree of control.

[3] Bénabou and Tirole (2003) formalise this crowding out under the assumption that the imposition of external rewards signals a low ability / high effort cost to the agent who is incompletely informed on this.

[4] Barkema (1995) provides evidence for this by showing that the effect of external intervention on work performance in Dutch firms is significantly positive (negative) in the case of impersonal (personal) control. Since intrinsic motivation tends to be more sensitive in personal relationships, the evidence lends some support to the crowding out hypothesis. Roomkin and Weisbrod (1999) provide evidence that managerial incentives are significantly weaker in non-profit as opposed to for-profit hospitals. This may be due to a number of explanations. First, it may reflect the consumption of rents as slack within non-profit organisations. Second, it may reflect that

Even in the absence of intrinsic motivation, agents may have a preference for autonomy. This is the case if the agent's performance and the degree of control are – at least to some extent – observable. In this case, a good performance under a greater degree of autonomy is clearly a better signal of an agent's ability than the same performance achieved under tight supervision.[5] Thus, the reputation and, with it, the prospective earnings of an agent tend to increase in the degree of autonomy.

When applying these ideas to the issue of budgetary delegation, one should expect the optimal mix of delegation and control to be determined not only by technological and informational considerations but also by its impact on the agents' motivation. Ignoring the latter could lead to distortions in the agents' effort, outweighing the benefits from greater control. More positively, the principal can use autonomy as an incentive device. It is the aim of the present analysis to address the implications for the allocation of budgets of agents' preferences for autonomy under moral hazard and adverse selection. We consider a model in which a principal allocates to an agent a budget for the purpose of production and determines the agent's autonomy in the use of it. She may also use a financial transfer to motivate the agent, where we distinguish the scenarios in which output is contractible and in which it is not. Agents differ in their efficiency in using a delegated budget either because they differ in ability or they face projects of different degrees of profitability. Here, we assume that an agent's type may be unknown to the principal. The agent's preferences over autonomy determine his effort incentives but also his preferences over different budgetary allocations. Thus, the level of autonomy granted to an agent is determined technologically by the agent's efficiency in handling the budget, motivationally by the agent's propensity to provide effort in return to autonomy, and by the agent's self-selection incentives in the presence of asymmetric information.

If agents value autonomy (control), the principal over- (under-)delegates (relative to the technological optimum excluding motivation) as a stimulus for the provision of effort. Under full information an efficient type receives both the greater budget and the greater degree of autonomy. Under asymmetric information, the principal distorts the overall budgets and the degree of delegation from their efficient levels in order to guarantee self-selection. In this, the agent's preference for autonomy turns out to play an important role. Unexpected allocations arise in the presence of a strong preference for autonomy. If the preference for autonomy is sufficiently strong but not too strong, the inefficient agent receives the greater budget and the degree of delegation is distorted downwards for the efficient agent and upwards for the inefficient agent. If the preference for autonomy is very strong the budget allocation depends on whether the principal employs a financial transfer as an additional instrument to generate self-selection. If this is the case, the efficient agent receives the greater budget but, perversely, a lower degree of delegation. The principal abstains from the use of financial incentives if the agent values the non-monetary benefit (job satisfaction) sufficiently more than the monetary benefit. In this case, the budgets are pooled if the preference for autonomy is strong. While we derive these results for a setting in which output is non-contractible, we demonstrate that they carry over to a setting in which output is contractible. Overall, our findings illustrate the potentially important role of the agents' preferences for autonomy, or preferences over the mode

non-profit hospitals perform additional tasks related to equity objectives or the provision of public goods. In as far as these tasks are difficult to monitor, the multi-task nature of the problem may require that weak incentives are implemented on all tasks. Finally, the absence of strong incentive systems is consistent with the presence of intrinsically motivated staff. This is particularly likely if self-selection occurs of intrinsically motivated managers into non-profit organisations.

[5] There is an extensive literature on the incentive effects of career concerns (e.g., Holmstrom 1999, Dewatripont et al. 1999), but to my knowledge it does not address the signal's dependence on the degree of autonomy.

of production more generally, for the purpose of intra-organisational resource allocation. This role becomes particularly pertinent under circumstances of asymmetric information and when the agent's preferences differ greatly from the principal's.

A recent literature deals with the effects on agency relationships of intrinsic motivation or, similarly, a public service spirit. Francois (2000) compares the incentives within public or non-profit as opposed to private providers when agents care in an altruistic way about the service provided.[6] Delfgaauw and Dur (2002) derive optimal incentive contracts for intrinsically motivated workers and consider the selection of workers into the firm when motivation is unobservable but may be signalled on the part of agents.[7] While asymmetric information about agent type pertains in our model as well, we consider screening rather than signalling. Glazer (2004) analyses how the principal chooses an input jointly with an intrinsically motivated agent's choice of effort, where inputs can be complements or substitutes and where moves can be simultaneous or sequential. In these models agents care about output but in contrast to our model their motivation does not depend on the mode of production.

Besley and Ghatak (2003) also model public service motivation but take into account that the organisation's mission influences incentives, where a mission is defined as the attributes of a project that make people value its success over and above any monetary rewards. Their analysis focuses on principal-agent matching rather than on optimal incentive schemes. Aghion and Tirole (1997) show that it pays principals to grant 'real' authority, even at the expense of control, if this induces agents to exert additional effort and relaxes their participation constraint.[8] Their analysis then focuses on how organisational overload or institutional arrangements commit principals to transfer real authority to agents even if they always retain formal authority. Murdock (2002) analyses how agents can be motivated by allowing them to establish a loss-making project with high intrinsic value to them in exchange for their effort towards a profit-making project. Due to the principal's lack of commitment these contracts will usually have to be relational rather than explicit. While these models share with ours the idea that the principal can enhance (intrinsic) motivation by giving up some control to the agent, they focus on the moral hazard problem alone and do not consider an adverse selection dimension that is central to our analysis.

Bénabou and Tirole (2003: section 3.1) consider the scope for crowding in intrinsic motivation by way of granting autonomy to an agent. They assume that, in contrast to the principal, the agent is only imperfectly informed about his type (high or low ability). By granting autonomy the principal not only provides a benefit to agents of all types, but she also signals that the agent's ability is high. By raising the agent's confidence this raises further the willingness to provide effort. Similar to our model, asymmetric information about the agent's type bears on the degree of delegation. However, the nature of this is rather different, as Bénabou and Tirole assume the principal rather than the agent to be the informed party. Finally, the context of all the aforementioned models is different in that they consider the expanse of effort on the development of profitable or otherwise projects, whereas we focus on the delegation of budgets.

---

[6] See also Dixit (2002). Heckman et al. (1996) provide empirical evidence supporting the presence of a public service spirit. They find that US social workers systematically select the least employable cases into job training programme in spite of performance incentives encouraging cream-skimming in favour of the most employable.

[7] Within different set-ups, Bénabou and Tirole (2003) and Grepperud and Pedersen (2004) derive optimal performance pay when intrinsic motivation may be crowded out.

[8] A similar spirit underlies the models by Gertner et al. (1994) and Mitusch (2000), where the agent's effort decreases in the degree of control as this delimits a real resource (rather than an intrinsic) rent.

In this latter aspect, our model is related to the principal-agent literature as applied to budgeting problems and is closest in spirit to Harris and Raviv (1998) and Bernardo et al. (2001).[9] Harris and Raviv (1998) consider an adverse selection setting without moral hazard in which the principal can use a costly audit as an instrument besides the budget assignments. They show that the optimal capital allocation generally implies over- (under-) investment for projects with low (high) productivity. The extent of this distortion increases in the audit cost. Harris and Raviv explain the scope for delegation – in the sense of the manager having a choice on capital allocation between two projects, which cannot be predicted by the principal – as an increasing function of the auditing cost. The focus of our analysis lies not so much with an explanation of delegation, the necessity of which we take as granted, but rather with the incentive role of delegation with regard to moral hazard and adverse selection. Similar to us, Bernardo et al. (2001) consider both moral hazard and adverse selection, where in both cases agents have an incentive to over-report profitability or level of ability in order to attract high budgets. As agents are not directly concerned about output in Bernardo et al., they can only be motivated by performance pay. While we also consider the scope for performance pay, one key instrument to stimulate effort is the budget allocation, as modified by the degree of autonomy.

The remainder of the paper is organised as follows. Section 2 introduces the model. Section 3 derives the full information optimum and discusses the role of delegation in resolving moral hazard when output is non-contractible. In section 4, we analyse in detail the allocation under asymmetric information maintaining the assumption of non-contractible output. Sections 5 and 6 deal with the case of contractible output in the absence and presence of asymmetric information, respectively. Section 7 provides a brief analysis of the case in which agents differ in their preference for autonomy and section 8 concludes. The proofs are relegated to an appendix.

2 The model

An agent produces an output or service, the value of which is given by a homogeneous function $H(e,b,r,\mathbf{b},\mathbf{r}) = eh(b,r,\mathbf{b},\mathbf{r})$, where e is a non-contractible effort, b is the delegated budget and r is the budget, which is retained under central control. The productivity of the budgetary inputs b and r is measured by $\mathbf{b}$ and $\mathbf{r}$, respectively. In order to simplify the analysis, we adopt a Cobb-Douglas specification

$$h(b,r,\mathbf{b},\mathbf{r}) = b^{\mathbf{b}} r^{\mathbf{r}}, \quad \mathbf{b} \in (0,1); \qquad \mathbf{r} \in (0,1) \qquad 2(\mathbf{b}_E + \mathbf{r}) < 1;^{10}$$

Thus, each budget exhibits non-negative but decreasing returns. A few words are warranted as to our interpretation of the delegated and retained budgets. Suppose the principal assigns an overall budget $B = b + r$ to the agent. The delegated budget b is the part of the overall allocation over the use of which the agent can dispose freely in the course of production. The retained budget r remains under the principal's control. It could either be spent by the principal on the purchase of inputs, which are then transferred to the agent for further use. Alternatively, the agent may have to obtain the principal's approval on the use of r, or only use this budget according to strict guidelines.

[9] For a more detailed review of this literature see Harris and Raviv (1998).

[10] The last constraint on the parameters guarantees concavity of the objective function.

Let us now introduce the concept of delegation and autonomy we have in mind. Define $D := \frac{b}{r}$, $D \in [0,\infty)$ as the 'degree of delegation'. It appeals to us to use the degree of delegation as a proxy measure of autonomy, understood as the absence of central intervention.[11] Using $D = \frac{b}{r}$ together with $B = b + r$ to rewrite $r = \frac{B}{1+D}$ and $b = \frac{DB}{1+D}$, we obtain the relationship $h(D,B,\cdot) = D^{b}\left(\frac{B}{1+D}\right)^{b+r}$ describing production as a function of the degree of delegation and the overall budget. Subsequently, we will make use of both specifications, $h(b,r,\cdot)$ and $h(D,B,\cdot)$ according to analytical and presentational convenience.

In using the function $h(b,r,b,r)$, the model remains general about the particular use in production of the budgetary inputs b and r. Likewise, we side-step the issue as to which activities should be delegated in order to focus on the incentive role of delegation. In this regard, the function $h(\cdot)$ is a 'black-box', which merely reflects that production can be organised in a variety of ways, involving a greater or lesser extent of delegation. The production elasticity b captures the efficiency of delegation. It can be understood to reflect the agent's technology and information as well as the transaction costs involved in delegation. For example, b may be low if the agent uses an inferior technology, if he has poorer access to information lower capacity of processing it, or if he has little bargaining power vis-à-vis local suppliers. Furthermore, b may reflect an agent's propensity to shirk. Alternatively, b may characterise the specific project to be carried out by the agent, where projects with a high b allow greater gains to delegation perhaps because they are of a non-standard nature so that there is little experience at the centre. The elasticity r captures the effectiveness of controlling budgets. It is determined by the cost of communicating relevant information to the centre as well as by the cost of communicating to the agents the centrally determined action plan. Furthermore, r depends on the cost of monitoring and enforcing the agents' compliance with the principal's plan.[12] It is readily verified that production $h(\cdot)$ is concave in D, implying the existence of a maximum at $\tilde{D} = \frac{b}{r}$. Indeed, if agents are unconcerned about autonomy, $\tilde{D}$ is the optimal degree of delegation. In the following, we will occasionally refer to it as the technologically efficient degree of delegation.

Each agent receives utility $V = U + at$, where U is the agent's non-monetary benefit of production and at is the value the agent places on a monetary transfer received from the principal, e.g. a salary. The agent's non-monetary benefit of production is given by

$$U = u(D)H(e,b,rb,r) - \frac{e^2}{2}; \quad D := \frac{b}{r} \tag{1}$$

It increases in output $H(\cdot)$ and decreases in the quadratic cost of effort. The extent to which an agent benefits from his production is governed by the weight $u(D)$, embracing the agent's preferences over autonomy and control. Specifically, let

---

[11] We define the degree of delegation as a factor-intensity for analytical convenience. A more intuitive measure of the degree of delegation would be the share of the delegated budget in the total budget $\frac{b}{b+r}$. It is easy to verify that for any two pairs (b,r) and (b',r') it is true that $\frac{b}{b+r} > \frac{b'}{b'+r'} \Leftrightarrow D = \frac{b}{r} > \frac{b'}{r'} = D'$. Thus, a greater degree of delegation implies and is implied by a greater share of the delegated budget in the total budget.

[12] In this regard, we distinguish the actions that can be controlled for by the principal from those for which the unverifiable effort e is relevant. The former could relate to spending decisions, whereas the latter would relate to the effort taken by the agent in drawing up the project and generating options that enhance its value.

$u(D) = kD^a$, $\quad a \in [-b, r]$, $\quad k > 0$,

where the parameter a reflects the agent's preference for autonomy. A greater degree of delegation, i.e. greater autonomy, raises the utility weight if and only if $a > 0$. The agent's concern about production could be explained by any of the following reasons. First, intrinsically motivated agents care about $H(\cdot)$, with the motivation increasing in the degree of autonomy. Second, agents may derive a 'warm glow' benefit à la Andreoni (1990) from providing to their customers a service $H(\cdot)$, a benefit they derive only to the extent of their 'personal' contribution towards it. Third, agents may be driven by professional status. As status usually rises with the degree of responsibility, it is plausible to assume that the agent's benefit from status increases not only with output but also with the extent to which this has been produced autonomously. Finally, $u(\cdot)H(\cdot)$ may be a measure of the agent's (discounted) future earnings, as determined by the reputation from having carried out the present task. It is reasonable to assume that the value of the reputation increases not only in the outcome but also in the degree of autonomy, as this measures the agent's individual as opposed to the organisation's contribution towards production. Note that in some instances agents may have a preference for central intervention such that $a < 0$. This may be due to a dislike for responsibility, when the agent suffers discomfort from a moral pressure to get the job right.

In the following, we assume that there are two types of agents/projects, efficient (E) and inefficient (I), where efficient agents/projects are characterised by $b_E > b_I$. Let $l \in [0,1]$ denote the probability of an agent/project being an E type, or alternatively the share of E types in the population. The principal levies budgets at a cost y, which we normalise to $y = 1$ without loss of generality. The (risk-neutral) principal's expected net value of production can then be written as

$$R(b_E, r_E, t_E, b_I, r_I, t_I) = \left\{ \begin{array}{l} l\left[H(e_E, b_E, r_E, b_E) - b_E - r_E - t_E\right] \\ +(1-l)\left[H(e_I, b_I, r_I, b_I) - b_I - r_I - t_I\right] \end{array} \right\} \qquad (2).$$

Sequence of moves

We follow the standard contracting framework under adverse selection as summarised in figure 1.

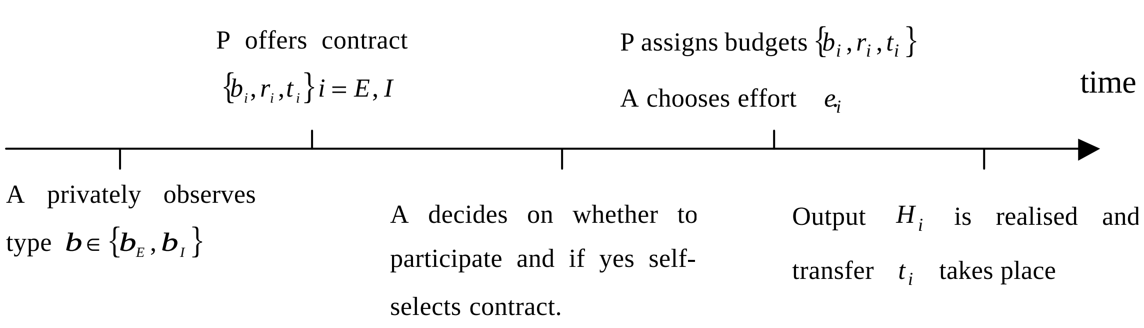

Figure 1: Timing of contract.

The agent privately observes his type $b \in \{b_I, b_E\}$. Then, the principal offers a contract $\{b_i, r_i, t_i\}$, which the agent accepts or rejects. Upon acceptance the agent self-selects. The contract is then executed with the principal providing the budgets and the agent choosing effort. Finally, output is realised and transfers take place. We will consider two scenarios.

Scenario 1 (sections 3 and 4). Neither effort e nor output H is contractible. Here, we have in mind a situation in which the agent does not produce a marketable output himself but rather an intermediate input into a more general production function. This embraces activities such as R&D, advertising or strategic planning, the contributions of which towards a company's profit are difficult to verify. Output verification is also difficult in case of a bureaucrat producing a non-market good or service. In this case, the transfer t_i from the principal to the agent cannot be made contingent on output but only on the budgetary allocation $\{b_i, r_i\}$. Note that the aforementioned activities often feature a strong role for motivation and autonomy.

Scenario 2 (sections 5 and 6). Here, effort continues to be non-contractible but output is now verifiable so that contracts can be written in which transfers depend on the realised output. This situation sometimes reflects the provision of expert services by members of the medical, educational and legal professions but also of more conventional management activities.[13]

Wealth and participation constraints

We assume that the agent faces a wealth constraint $t \geq t_0$, where we normalise $t_0 = 0$ without loss of generality. More importantly, we assume that the agent's outside benefit V_0 is such that participation at $t \geq 0$ is always guaranteed. For the sake of simplicity, assume $V_0 = 0$. In this case, the requirement of non-negative transfers and the non-verifiability of effort imply that the agent's ex-post utility satisfies $V = U + at \geq U \geq 0$, where the second inequality follows from the fact that the agent could always choose $e = 0$. Hence, participation is guaranteed irrespective of the agent's type. Generally, the interaction of the wealth and participation constraints defines a range of different regimes similar to those featured in Laffont and Martimort (2002: section 3.5). With the main interest of the paper lying on the regime with a slack participation constraint, we will presently focus on this case without providing a complete characterisation. At this stage, merely note that a slack participation constraint is not unlikely in situations in which the agent's non-monetary benefit is substantial such as in the professional and/or 'creative' services we have in mind. In particular, if these services are built on longer-term relationships or involve specific investments, the agent's non-monetary benefit may lie well above his outside utility so that rent equalisation would require a payment from the agent to the principal. We assume that such payments are ruled out by wealth constraints, or in other words, that the agent's non-monetary rents cannot be (fully) extracted.

3 Allocation under full information when output is not contractible: Motivation by delegation

When neither effort nor output are contractible (scenario 1) the principal cannot use the transfers to stimulate effort on the part of the agent. In this case, it is optimal to set $t_E = t_I = 0$,

[13] Note that in many cases, outcome related pay is not instituted even within the professions. This is the case where outcomes are hard to verify (health as the outcome of medical services) and/or where the professional works as part of a team (education).

where the wealth constraint binds. Under complete information, the principal can identify the type of individual agents/projects and accurately allocates type specific budgets so as to $\max_{b_E, r_E, b_I, r_I} R\big|_{t_E = t_I = 0}$ as given in (2). In so doing, she will take into account the effect of delegation on the agent's motivation and provision of effort. Solving the problem backwards, we begin by considering the agent's choice of effort. Given the budget, the agent chooses effort so as to maximise utility (1). From the first-order condition

$$e^* = \hat{e}(b, r, \boldsymbol{b}) = u(D)h(b, r, \boldsymbol{b}) = kb^{b+a} r^{r-a} \tag{3),}$$

we obtain output and utility as functions of the budgets only

$$\hat{H}(b, r, \boldsymbol{b}) = \hat{e}(b, r, \boldsymbol{b})h(b, r, \boldsymbol{b}) = u(D)h(b, r, \boldsymbol{b})^2 = kb^{2b+a} r^{2r-a} \tag{4),}$$

$$\hat{U}(b, r, \boldsymbol{b}) = \hat{e}(b, r, \boldsymbol{b})\left[u(D)h(b, r, \boldsymbol{b}) - \tfrac{1}{2}\hat{e}(b, r, \boldsymbol{b})\right] = \tfrac{1}{2}\left[u(D)h(b, r, \boldsymbol{b})\right]^2 = \tfrac{1}{2}k^2 b^{2(b+a)} r^{2(r-a)} \tag{5).}$$

Placing the restrictions $b \geq 1$ and $r \geq 1$, we obtain $\hat{H}_b(b, r, \boldsymbol{b}) > 0$ and $\hat{U}_b(b, r, \boldsymbol{b}) > 0$ as well as $\hat{H}_{xb}(b, r, \boldsymbol{b}) > 0$ and $\hat{U}_{xb}(b, r, \boldsymbol{b}) > 0$, $x = b, r$. Hence, total and marginal output as well as total and marginal utility are greater for the E type, reflecting the greater productivity. Again, we will sometimes express effort, output and utility as functions $\hat{e}(B, D, \boldsymbol{b})$, $\hat{H}(B, D, \boldsymbol{b})$ and $\hat{U}(B, D, \boldsymbol{b})$. One can then easily check that equilibrium effort and utility are concave in D, with $\operatorname{sgn} \frac{\partial \hat{e}}{\partial D} = \operatorname{sgn} \frac{\partial \hat{U}}{\partial D} = \operatorname{sgn}\left[\boldsymbol{a} + \boldsymbol{b} - (\boldsymbol{r} - \boldsymbol{a})D\right]$. The principal can then stimulate extra effort by increasing the degree of delegation up to the level

$$\hat{D}_i = \frac{\boldsymbol{b}_i + \boldsymbol{a}}{\boldsymbol{r} - \boldsymbol{a}} = \arg\max \hat{U}(B_i, D_i, \boldsymbol{b}_i); \quad i = E, I \tag{6}$$

that maximises the agent's utility. The concavity of $\hat{e}(B, D, \boldsymbol{b})$ in D reflects the conventional wisdom that granting some responsibility tends to make people work harder, while too much delegation encourages slack.

Consider now the principal's choice of budgets. Using (4) in (2) we obtain the first-order conditions

$$\hat{H}_b(b_i, r_i, \boldsymbol{b}_i) - 1 = 0 \quad \text{(7a)} \qquad \hat{H}_r(b_i, r_i, \boldsymbol{b}_i) - 1 = 0; \quad i = E, I \quad \text{(7b).}$$[14]

The full information optimum $\{b_E^*, r_E^*, b_I^*, r_I^*\}$ can then be characterised as follows.

Proposition 1. (i) The efficient agent receives both a greater delegated and a greater controlled budget, i.e. $b_E^* > b_I^*$ *and* $r_E^* > r_I^*$. *(ii) Moral hazard leads to an upward (downward) distortion in the degree of delegation for both types if and only if agents have a preference for*

---

[14] The second-order condition holds if and only if $\hat{H}_{bb}(\cdot, \boldsymbol{b}_i)\hat{H}_{rr}(\cdot, \boldsymbol{b}_i) - \left[\hat{H}_{br}(\cdot, \boldsymbol{b}_i)\right]^2 > 0; \quad i = E, I$. Using (4) and observing $2(\boldsymbol{b}_E + \boldsymbol{r}) < 1$ and $\boldsymbol{b}_E > \boldsymbol{b}_I$, it is readily checked that this is satisfied.

(against) autonomy, i.e. if and only if $a > 0$ *(*$a < 0$*). (iii) The principal delegates to greater extent to the efficient agent, i.e.* $D_E^* > D_I^*$.

Proof: See Appendix.

The principal chooses the delegated and controlled budgets not only with a view to 'technological' efficiency but also with a view to eliciting effort. The optimal degree of delegation under moral hazard can be written as

$$D_i^* = \tilde{D}_i + \overbrace{h(\cdot, b_i)\hat{e}_r(\cdot, b_i)}^{>0}\left(\hat{D}_i - \tilde{D}_i\right) = \frac{(2b_i + a)}{2r - a}; \quad i = E, I \tag{8}$$

where $\tilde{D}_i = \frac{b_i}{r}$ is the technologically efficient degree of delegation and where $\hat{D}_i$, as defined in (6), is the degree of delegation preferred by the agent. Here, $\hat{D}_i > \tilde{D}_i \Leftrightarrow a > 0$ implies that the agent prefers a degree of delegation over and above the technologically efficient level if and only if he has a preference for autonomy. In this case, it is optimal for the principal to over-delegate as this stimulates additional effort.

As an agent's utility increases both in b and r, we have $\hat{U}\left(b_E^*, r_E^*, b_i\right) > \hat{U}\left(b_I^*, r_I^*, b_i\right)$ and both types prefer the budget allocated to the E type.[15] While the principal can impose the budgets under full information, inefficient agents have an incentive to misrepresent their type under asymmetric information, thereby causing a problem of adverse selection.[16]

4 Allocation under asymmetric information when output is not contractible: Screening by delegation

From now on, we consider the level of efficiency b to be an agent's private information, the principal only being informed about the distribution of types. The principal's problem is now to $\max_{b_i, r_i, t_i;\, i=E,I} R$, as given in (2), subject to the self-selection constraints

$$\hat{U}\left(b_I, r_I, b_I\right) + at_I \geq \hat{U}\left(b_E, r_E, b_I\right) + at_E \tag{ICI}$$

$$\hat{U}\left(b_E, r_E, b_E\right) + at_E \geq \hat{U}\left(b_I, r_I, b_E\right) + at_I, \tag{ICE}$$

and the wealth constraints $t_E \geq 0$ and $t_I \geq 0$. Let m_1 and m_2 denote the multipliers associated with the (ICI) and (ICE) constraint, respectively, and let m_3 and m_4 denote the multipliers associated with the wealth constraints $t_I \geq 0$ and $t_E \geq 0$, respectively. We can then write the first-order conditions associated with t_I, t_E, b_I, r_I, b_E and r_E as follows.

[15] In this, our model is similar to Harris and Raviv (1998). It is different, however, from the standard model (Laffont and Martimort 2002: section 2.3), where the efficient agent has an incentive to mimic the inefficient one. The difference arises from the principal's inability to fully extract rents in our model, on the one hand, and the assumption of a common reservation utility in the standard model, on the other hand.

[16] Note that a situation of natural separation may be possible if $b_E^* > b_I^*$ and $r_E^* < r_I^*$. This is ruled out for Cobb-Douglas-preferences but may be possible for more general specifications of $\hat{U}$.

$$-(1-l)+a(m_1-m_2)+m_3=0 \qquad (9a),$$

$$-l-a(m_1-m_2)+m_4=0 \qquad (9b),$$

$$(1-l)\left[\hat{H}_b(b_I,r_I,\boldsymbol{b}_I)-1\right]+m_1\hat{U}_b(b_I,r_I,\boldsymbol{b}_I)-m_2\hat{U}_b(b_I,r_I,\boldsymbol{b}_E)=0 \qquad (9c),$$

$$(1-l)\left[\hat{H}_r(b_I,r_I,\boldsymbol{b}_I)-1\right]+m_1\hat{U}_r(b_I,r_I,\boldsymbol{b}_I)-m_2\hat{U}_r(b_I,r_I,\boldsymbol{b}_E)=0 \qquad (9d),$$

$$l\left[\hat{H}_b(b_E,r_E,\boldsymbol{b}_E)-1\right]-m_1\hat{U}_b(b_E,r_E,\boldsymbol{b}_I)+m_2\hat{U}_b(b_E,r_E,\boldsymbol{b}_E)=0 \qquad (9e),$$

$$l\left[\hat{H}_r(b_E,r_E,\boldsymbol{b}_E)-1\right]-m_1\hat{U}_r(b_E,r_E,\boldsymbol{b}_I)+m_2\hat{U}_r(b_E,r_E,\boldsymbol{b}_E)=0 \qquad (9f).[17]$$

Let $\{b_E^{**},r_E^{**},t_E^{**},b_I^{**},r_I^{**},t_I^{**}\}$ denote the optimal allocation. The following Lemma will prove helpful in understanding the structure of the shadow prices.

Lemma 1. The optimum entails (i) $m_1>m_2\geq 0$; (ii) $m_4>0$; (iii) $m_2=0$ if $D_E>D_I$ or if $D_E=D_I$ and $B_E\neq B_I$; and (iv) $m_3=0$ if $D_E<D_I$ or if $D_E=D_I$ and $B_E\neq B_I$.

Proof: See Appendix.

According to part (i) of the Lemma an optimum always entails a binding (ICI). While (ICE) will be slack in many cases, it cannot be ruled out that (ICI) and (ICE) bind simultaneously. Part (ii) of the Lemma implies that it is always optimal to set $t_E^{**}=0$. Part (iii) shows that (ICE) binds only if either the degree of delegation has been reversed from the full information situation such that $D_I>D_E$ or if a full pooling equilibrium is realised. Likewise, according to part (iv) the transfer to the I type must be positive, i.e. $t_I^{**}>0$, either if the degree of delegation is reversed or if the degree of delegation is pooled but not total budgets.

Agents are not motivated by monetary transfers: $a=0$

Consider now the special case $a=0$, where agents do not respond to financial incentives and are merely motivated by the job. Obviously, it is then optimal for the principal to set $t_E^{**}=t_I^{**}=0$. Although this case may lack intuitive appeal we consider it in order to illustrate a set of first results.[18] The same results apply for a setting with $a>0$, as long as a is sufficiently low, i.e. as long as financial rewards are not too important in the agent's utility. We can then characterise as follows the degree of delegation assigned to each type in a separating equilibrium.

[17] The second-order conditions are satisfied for $2(\boldsymbol{b}_E+\boldsymbol{r})<1$ if the parameter k in the function $\hat{U}(\cdot,k)$ is sufficiently low.

[18] One reason for $a=0$ may lie in a lexicographic preference ordering. Assume that instead of $V=U+at$, the agents' preferences are given by $\hat{V}(t,U)$, where $t>t' \Leftrightarrow \hat{V}(t,U)>\hat{V}(t',U') \quad \forall U,U'$. In this case, all agents strictly prefer a greater salary irrespective of the budgetary allocation. In this case, the principal can attain separation only if she sets $t_E=t_I$, in case of which the monetary transfer becomes irrelevant for incentive purposes and $t_E=t_I=0$.

Lemma 2. The optimal levels of delegation D_I^{**} *and* D_E^{**} *are given by*

$$D_I^{**} = D_I^{*} + \frac{m_1 \hat{U}_r(b_I, r_I, b_I)}{1-l}\left(\hat{D}_I - D_I^{*}\right) + \frac{m_2 \hat{U}_r(b_I, r_I, b_E)}{1-l}\left(D_I^{*} - \hat{D}_E\right) \quad \text{(10a);}$$

$$D_E^{**} = D_E^{*} + \frac{m_1 \hat{U}_r(b_E, r_E, b_I)}{l}\left(D_E^{*} - \hat{D}_I\right) + \frac{m_2 \hat{U}_r(b_E, r_E, b_E)}{l}\left(\hat{D}_E - D_E^{*}\right) \quad \text{(10b).}$$

Proof: See Appendix.

Generally, the optimal degree of delegation under asymmetric information deviates from the first-best levels with the direction of the deviation being determined by the sign of the 2nd and 3rd terms in (10a) and (10b). Here, (ICI) requires an increase (decrease) in the I type's degree of delegation if the agent prefers a degree of delegation, $\hat{D}_I$, that exceeds (falls short of) the degree of delegation D_I^{*}, the principal would assign to this type under full information. Likewise, (ICI) requires an increase (decrease) in the E type's degree of delegation if $\hat{D}_I$ falls short of (exceeds) the degree of delegation D_E^{*} the principal would optimally assign to the E type. Recall from Lemma 1 that $m_1 > 0$ is always true whereas $m_2 > 0$ requires $D_I^{**} > D_E^{**}$. We define

$$\bar{a} := \frac{2(b_E - b_I)r}{r + 2b_E - b_I} \quad \text{(11),}$$

and

$$\hat{a} := \frac{2(b_E - b_I) r b_P^{\,2b_E}}{(r + 2b_E - b_I) b_P^{\,2b_E} - (r + b_I) b_P^{\,2b_I}} \quad \text{(12),}$$

where $b_P = b_E = b_I$ is the pooling level of the controlled budget. Note that $\hat{a} \in \left]\bar{a}, r\right[$. We can now distinguish four regimes.

Lemma 3. (i) $a < 0 \Leftrightarrow D_I^{**} < D_I^{*} < D_E^{*} < D_E^{**}$; *(ii)* $a \in \left[0, \bar{a}\right[\Leftrightarrow D_I^{*} \le D_I^{**} < D_E^{*} < D_E^{**}$; *(iii)* $a \in \left[\bar{a}, \hat{a}\right[\Leftrightarrow D_I^{*} < D_I^{**} < D_E^{**} \le D_E^{*}$; *(iv)* $a \ge \hat{a} \Leftrightarrow D_E^{**} = D_I^{**} = D^P$, *where* D^P *is the pooling degree of delegation.*

Proof: See Appendix.

We thus know how the degree of delegation for each type evolves with the preference for autonomy a. Before we discuss the different separating and the pooling allocation more fully let us establish how the total budgets B_E^{**} and B_I^{**} develop with a.

Lemma 4. (i) $a < 0 \Leftrightarrow B_I^{*} < B_I^{**} < B_E^{**} < B_E^{*}$; *(ii)* $a \in \left[\bar{a}, \hat{a}\right[\Leftrightarrow \max\left\{B_I^{*}, B_E^{**}\right\} < \min\left\{B_I^{**}, B_E^{*}\right\}$; *(iii)* $a \ge \hat{a} \Leftrightarrow B_I^{*} < B_I^{**} = B^P = B_E^{**} < B_E^{*}$, *where* B^P *is the pooled total budget.*

Proof: See Appendix.

We observe immediately that the principal always under-budgets (over-budgets) the efficient (inefficient) type relative to the first-best. Generally, self-selection requires that a rent is paid to the informed agent, in our case the I type. This always involves an increase in the overall budget paid to the I type. In this regard, our model is similar to the standard model of agency (Laffont and Martimort 2002: section 2.6). In contrast to the standard case, however, rental payments are directly associated with an efficiency loss due to the unproductive use of funds, as can be seen from the first-order conditions (9e) and (9f), where the marginal product of b_I and r_I, respectively, falls short of the marginal cost of funds. By reducing the E type's budget from its full information level, the principal lowers the attractiveness of mimicking this type and is, thus, able to reduce the informational rent to the I type. In so doing, she trades-off the efficiency losses for the two types. This resembles the finding by Harris and Raviv (1998), where in the presence of an imperfect monitoring technology, the principal reacts to a situation of asymmetric information by reducing the gap in the budgets aimed at the efficient and inefficient agent. Bearing this in mind, we can now focus on a number of more salient aspects of the allocation under asymmetric information.

Proposition 2. (a) If agents have a preference for control ($a < 0$), the principal over (under-) delegates to the efficient (inefficient) type relative to the first-best. (b) If agents have a weak preference for autonomy ($a \in [0, \overline{a}[$), the principal over-delegates to both types. She pays a greater total budget to the inefficient type if and only if the degree of delegation is sufficiently high. (c) If agents have a strong preference for autonomy ($a \in [\overline{a}, \hat{a}[$), the principal under (over-) delegates to the efficient (inefficient) type. She pays a greater total budget to the inefficient type. (d) If agents have a very strong preference for autonomy ($a \in [\hat{a}, r]$), the principal pools the degree of delegation and the budgets.

According to the agents' preference for autonomy, we can, thus, distinguish four regimes, three involving separation, (a)-(c), and one involving pooling, (d). The principal uses the degree of delegation as an additional instrument in separating types. This is best illustrated with reference to the benchmark case in which agents do not have preferences about the mode of production, i.e., the degree of autonomy. In this case, $a = 0$ which from (10a) and (10b) implies $D_E^{**} > D_E^{*} > \hat{D}_I$ and $D_I^{**} = D_I^{*} = \hat{D}_I$. Since the I type prefers the same degree of delegation as the principal, any deviation from the optimal level would only lead to a reduction in job satisfaction on the I type's own contract and thus to an increase in the information rent. In contrast, an increase in the E type's degree of delegation renders this allocation less attractive to the I type and, thereby, helps to contain the information rent and the associated distortions in the budgets B_I^{**} and B_E^{**}. Here, the principal introduces a second-order loss of efficiency into the E type's delegation in order to achieve a first-order gain in efficiency with regard to the levels of budgets.

If the agent's preference for autonomy deviates from the principal's, i.e. if $a \neq 0$, the principal introduces a distortion into the inefficient type's degree of autonomy, $D_I^{**} \neq D_I^{*}$. The reason is that the principal and the agent no longer agree on the optimal degree of delegation. There is now scope for the principal to pay out a part of the informational rent by granting the I type a more preferred mode of production and, thereby, reduce the rent paid in real resources as well as the distortions in the E type's allocation. If agents dislike autonomy, i.e. if $a < 0$ [regime (a)], the agent's preferred degree of autonomy $\hat{D}_I$ lies below the one preferred by the principal D_I^*. The principal can then enhance the I type's utility on the own contract by reducing the degree of delegation to a level $D_I^{**} \in \left[\hat{D}_I, D_I^*\right]$. Over-delegation to the E type i.e. $D_E^{**} > D_E^* > D_I^*$ makes this allocation even less attractive, and, thus, allows a further reduction in the budgetary distortion.

If the preference a is positive but sufficiently low relative to the productivity spread $\boldsymbol{b}_E - \boldsymbol{b}_I$ [regime (b)], the I type's preferred degree of delegation lies between the first-best for the E and I type respectively, i.e. $D_I^* < \hat{D}_I < D_E^*$. It is then best for the principal to over-delegate to the I type in order to contain the total budget B_I. But as the productivity spread is large, the E type's optimal degree of delegation is still unattractive for the I type and over-delegation to E continues to be optimal. If a is large relative to the productivity spread [regime (c)], gaining greater autonomy becomes a very strong incentive for I when seeking to select E's allocation. In fact, $\hat{D}_I > D_E^*$ so that the I type prefers a degree of delegation in excess of the E type's first-best. While still over-delegating to I, the principal now optimally reduces the degree of delegation to the E type below the full information level. While E types still receive the greater degree of delegation, they now receive a total budget below the one assigned to I types. This is illustrated in figure 2 a), with the budgetary assignments E to the E type and E' to the I type. The dashed line $B_E B_E$ gives the iso-budget curve corresponding to B_E^{**}, with $B_E^{**} < B_I^{**}$. With both E and E' lying on the same indifference curve $\hat{U}_I$, the I type is just indifferent. Being more efficient under a greater degree of delegation, the E type strictly prefers the allocation at E despite the lower overall budget. As total budgets are reversed, the distortion goes beyond the rationing of capital for efficient types and the over-funding of inefficient types as e.g. in Harris and Raviv (1998). Greater autonomy is now effectively traded against a lower total budget.

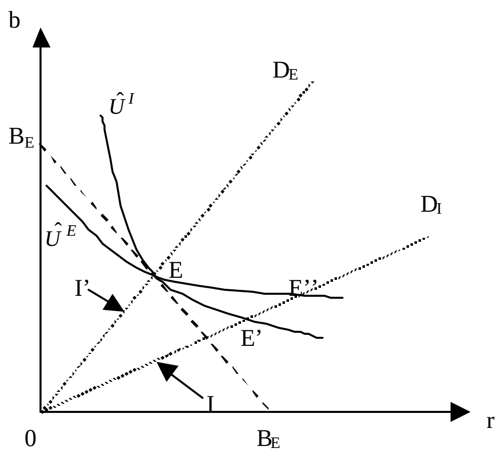

Figure 2a: Separation in regime (b).

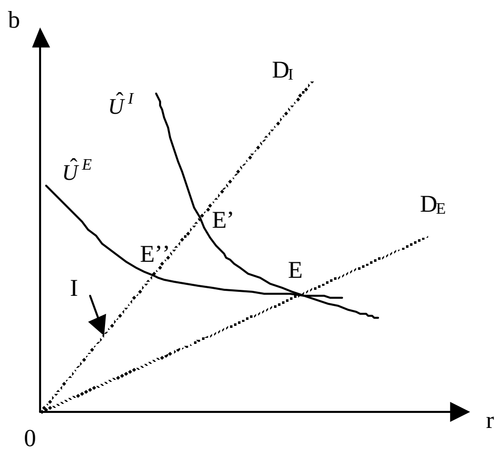

Figure 2b: Non-existence of separation in regime (d).

Finally, for a very high preference for autonomy only a pooling allocation is feasible, whereby both agents receive the same budget and the same degree of delegation [regime (d)]. Here, the agent's preference for autonomy is so strong that the principal would effectively have to assign a degree of delegation to the I type greater than the one assigned to the E type. Figure 2 b) illustrates that this is impossible. Suppose E and E' are assigned to the E and I type, respectively, such that $D_E^{**} < D_I^{**}$. While the I type is indifferent, this allocation violates the E type’s incentive constraint (ICE). While a reduction in the overall budget to the level at E’’ would restore incentive compatibility for the E type, it is now violated for the I type. The best the principal can then attain is an allocation in which both budgets and the degree of delegation are pooled.

From (ICI) and (ICE) one can derive the monotonicity condition

$$\hat{U}_{Bb}(\cdot, b_I)\Delta B + \hat{U}_{Db}(\cdot, b_I)\Delta D \geq 0 \qquad (13),$$

with $\Delta B = B_E - B_I$ and $\Delta D = D_E - D_I$, that is necessary for the existence of a separating allocation.[19] Satisfaction of (ICI) requires $\Delta B = \frac{-U_D(\cdot, b_I)}{U_B(\cdot, b_I)}\Delta D$. Inserting this into (13) yields the equivalent condition $\left[\frac{-U_D(\cdot, b_I)}{U_B(\cdot, b_I)}\hat{U}_{Bb}(\cdot, b_I) + \hat{U}_{Db}(\cdot, b_I)\right]\Delta D \geq 0$. Inserting the appropriate derivatives from (5) one can rewrite the condition to $\frac{4[\hat{U}(\cdot, b_I)]^2 (r-a)}{D_I B_I}\Delta D \geq 0 \Leftrightarrow \Delta D \geq 0$. This condition is violated in regime (d), where $a > \hat{a}$ would imply $\Delta D < 0$. The underlying reason is a direct conflict between the incentive compatibility constraints (ICI) and (ICE) that cannot be resolved in the absence of transfers. Guesnerie and Laffont (1984) call an environment in which separation becomes unfeasible ‘non-responsive’. In their set-up and, similarly, in Laffont and Martimort (2002: section 2.10.2) non-responsiveness occurs as a result from a conflict between the allocation that maximises total surplus and the monotonicity condition. In our case, this is different. It is easily checked that the maximisation of total surplus $l\left[\hat{H}(\cdot, b_E) + \hat{U}(\cdot, b_E)\right] + (1-l)\left[\hat{H}(\cdot, b_I) + \hat{U}(\cdot, b_I)\right]$ implies $\Delta D > 0$ and is, therefore, compatible with monotonicity. Non-responsiveness occurs due to a conflict between rent extraction (from the I type), requiring $\Delta D < 0$ and the monotonicity condition. In this case, separation would increase rental payments by more than it would increase the value of production.[20]

Agents are motivated by monetary transfers: $a > 0$

Let us now characterise the allocation for the case $a > 0$, where monetary transfers become available to the principal as an additional tool to motivate agents. Specifically, the principal is now able to pay out some of the I type’s informational rent by way of a monetary transfer $t_I^{**} > 0$. Nonetheless, we will see that there are still cases in which the principal prefers not to use monetary transfers at all. In (a, a) space we can identify five regimes for the budgetary allocation corresponding to the areas I-V in figure 3, as developed in Lemma 5 and Corollary L5. For the purpose of this illustration we focus on $a \geq 0$ without loss of generality.

---

[19] Here, we express utility in terms of total budget and degree of delegation, $\hat{U}(B, D, b)$. The appropriate function and derivatives are easily determined from (5) when setting $b = \frac{DB}{1+D}$ and $r = \frac{B}{1+D}$.

[20] Morand and Thomas (2003) provide, within a different set-up, conditions for such a clash between rent extraction and monotonicity.

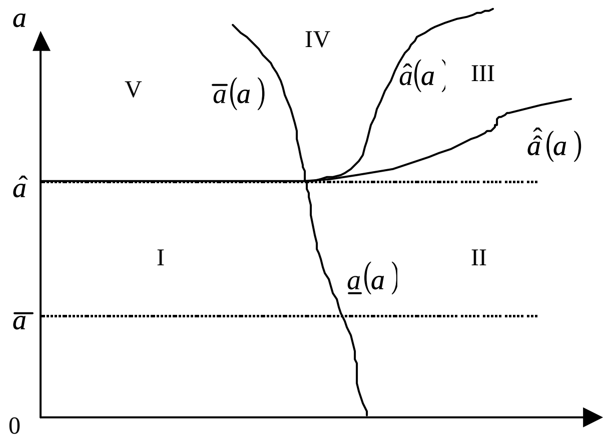

Figure 3: Budgetary regimes in (a, α) space.

Lemma 5. (i) Consider $\alpha < \hat{\alpha}$. *There exists a function* $\underline{a}(\alpha) \in (0, \infty)$ *such that* $m_3 > 0 \Leftrightarrow a \leq \underline{a}(\alpha)$. *(ii) Consider* $\alpha \geq \hat{\alpha}$. *There exists a unique function* $\overline{a}(\alpha) \in (0, \infty)$ *such that* $m_3 > 0 \Leftrightarrow a \leq \overline{a}(\alpha)$*; a unique correspondence* $\hat{a}(\alpha) \in (0, \infty)$ *such that* $m_2 > 0 \Leftrightarrow a \leq \hat{a}(\alpha)$*; and a unique function* $\hat{\hat{a}}(\alpha) \in (0, \infty)$ *such that* $D_I^{**} \geq D_E^{**} \Leftrightarrow a \leq \hat{\hat{a}}(\alpha)$. *(iii)* $\overline{a}(\alpha) \leq \hat{a}(\alpha) \leq \hat{\hat{a}}(\alpha)$ *and* $\underline{a}(\hat{\alpha}) = \overline{a}(\hat{\alpha}) = \hat{a}(\hat{\alpha}) = \hat{\hat{a}}(\hat{\alpha})$.

Proof: See Appendix.

Remark: The negative slope of the $\underline{a}(\alpha)$ and $\overline{a}(\alpha)$ schedules, as well as the positive slope of the $\hat{\hat{a}}(\alpha)$ schedule can be established from the system (9a)-(9f) under use of the implicit function theorem.

Corollary L5: *(i)* $t_I \begin{Bmatrix} = 0 \\ > 0 \end{Bmatrix}$ in $\begin{cases} \text{areas I and V} \\ \text{areas II, III and IV} \end{cases}$;

(ii) $D_E^{**} \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} D_I^{**}$ in $\begin{cases} \text{areas I and II} \\ \text{area V} \\ \text{areas III and IV} \end{cases}$.

Proof: See Appendix.

Area I: Here $t_I^{**} = 0$ and the allocation $\{B_E^{**}, D_E^{**}, B_I^{**}, D_I^{**}\}$ is, therefore, identical to the one realised for $\{a = 0;\ \alpha < \hat{\alpha}\}$ as characterised in Proposition 2. Monetary transfers are not used as their cost (in real resource terms) to the principal, $\frac{(1-l)}{a}$, exceeds the cost of an inefficient budget allocation as captured by the shadow price m_1. This is the case if a weak preference for money (a low a) on the agent's part would require the payment of substantial transfers in order to improve the efficiency of the budget allocation. The negative slope of the $\underline{a}(\alpha)$

schedule indicates that for $\alpha > 0$ financial transfers are the less likely to be used the lower the agents' preference for autonomy. This is because the inefficiency in the budget allocation falls as α decreases towards zero, where the agent's and principal's preferences with regard to autonomy converge.

Area II: Here $t_I^{**} > 0$, where the cost of monetary transfers is sufficiently low to render efficient the use of the monetary transfer as an additional screening instrument. This allows a reduction in the screening-induced inefficiencies in budgeting. Specifically, for a higher α the principal increases B_E^{**} (lowers B_I^{**}), and, thereby, mitigates capital rationing of efficient types (over-budgeting of inefficient types). A greater α also allows reducing the distortions in the degree of autonomy allocated to each type. For instance, an increase in α allows the principal to mitigate the under- (over-) delegation towards the efficient (inefficient) type that arises for $a > \overline{a}$.

Areas III and IV: Here, $t_I^{**} > 0$ and $D_E^{**} < D_I^{**}$, implying a reversal of the degree of delegation. The E type receives the greater budget, $B_E^{**} > B_I^{**}$ which in area III is sufficient to guarantee a slack incentive constraint (ICE) for the E type. Nonetheless, the budgetary allocation is perverse in that the principal assigns the greater degree of delegation to the I type despite its lower productivity in the use of the delegated budget.[21] This distortion increases as the preference for monetary rewards α falls. A lower effectiveness of monetary transfers in inducing self-selection on the part of the I type also implies that the principal has to close the gap $B_E^{**} - B_I^{**} > 0$. As the E type's allocation is thus rendered less and less attractive, this leads to a situation in area IV where the incentive constraints bind for both types.

Area V: Here $t_I^{**} = 0$, where the preference for monetary rewards, α, has dropped by so much that it is no longer efficient for the principal to use transfers. While this regime is akin to regime I in this regard, the high preference for autonomy rules out a separating allocation. Thus, the budget allocation is fully pooled, where $D_E^{**} = D_I^{**}$ and $B_E^{**} = B_I^{**}$. The negatively sloped $\overline{a}(\alpha)$ schedule implies that financial incentives are more likely to be used - and pooling is thus the less likely - the greater the agents' preference for autonomy and, therefore, the greater the distortions in the budget allocation. We can summarise as follows.

Proposition 3. (i) The principal uses financial incentives if and only if the agent's preference for financial rewards over job-satisfaction is sufficiently high. (ii) Budgets are fully pooled if the preference for financial rewards is sufficiently low, and the preference for autonomy is sufficiently high but not too high. (iii) The degree of delegation is reversed if the preference for autonomy is sufficiently high and the preference for financial rewards is at an intermediate level.

The principal will not always use financial instruments to motivate agents' self-selection even if they are available. This applies even if agents have a preference for financial rewards, $\alpha > 0$, as long as it is weak relative to their preference over other job attributes – in our case the satisfaction from a job done under some preferred degree of autonomy. It is then less costly for the principal to distort the budget allocation rather than to use financial rewards. As the preference for autonomy grows, the principal can induce separation only by introducing increasingly stronger distortions into the budgets. Specifically, she will have to render more

[21] This notwithstanding the efficient type still receives the greater delegated budget, i.e. $b_E^{**} > b_I^{**}$.

similar the degree of delegation. Full pooling of budgets emerges if rent extraction leads to a conflict between the two types' incentive constraints. Notably, pooling arises not only as a corner solution for $a = 0$, but it is supported by a range of pairs (a, a) in region V. Although the principal could break a pooling equilibrium by use of financial transfers, she refrains from this when agents are disinterested in financial rewards and their motivation requires inefficiently high payments.[22]

Why separation is feasible within regime IV (and similarly within regime III) can be understood as follows. Both incentive constraints (ICE) and (ICI) continue to bind within regime IV, as they did within the pooling regime V. In contrast to regime V, however, the efficient use of financial transfers allows the principal to reconcile the conflicting self-selection constraints. This is illustrated in figure 2b. Consider the budget allocations E to the E type and I to the I type, where $\Delta D = D_E - D_I < 0$ and $\Delta B = B_E - B_I > 0$. While allocation E is preferred by both types, the principal sets the transfer $t_I^{**} > 0$ such that (ICI) is just binding. (ICE) is then satisfied if the monotonicity condition (13) holds. Satisfaction of (ICI) requires $\Delta B = \frac{-U_D(\cdot, b_I)}{U_B(\cdot, b_I)} \Delta D + \frac{a t_I}{U_B(\cdot, b_I)}$. Inserting into (13) first this and then the appropriate derivatives from (5) yields the equivalent condition $\frac{4\hat{U}(\cdot, b_I)^2 (r - a)}{D_I B_I} \Delta D + \frac{a t_I}{U_B(\cdot, b_I)} \geq 0$. Since $a > 0$ allocations involving $\Delta D < 0$ are now feasible. Here, the spread in budgets $\Delta B > 0$ can be made sufficiently large to induce the E type to self-select despite of a lower degree of delegation and a lower financial transfer.

5 Allocation under full information when output is contractible

In this section we study the case in which output is verifiable (scenario 2), which allows the principal to link the agent's compensation to output and not merely to the budget allocation. As before, we establish as a benchmark the allocation when the principal can observe the agent's type b but not effort (i.e. moral hazard only), before moving on to the case in which the principal is uninformed about the agent's type b and effort (i.e. adverse selection cum moral hazard). When output is contractible the principal pays a transfer t_i to agent $i = E, I$ if and only if the agent produces a pre-specified output level $\overline{H}_i$ such that

$$t_i \begin{cases} > 0 \Leftrightarrow H(e_i, b_i, r_i, b_i) = \overline{H}_i \\ = 0 \Leftrightarrow H(e_i, b_i, r_i, b_i) \neq \overline{H}_i \end{cases}.$$

Note that under the wealth constraint the principal cannot punish the agent for deviations from the target. As there is no uncertainty involved in production we are dealing with a case of 'false moral hazard' (e.g. Laffont and Martimort 2002: section 7.1.4). Given the budget allocation $\{b_i, r_i\}$ a production target $\overline{H}_i$ implies an effort $\overline{e}_i$.[23] We can then write the moral hazard constraint for type i as

[22] While we have derived this finding for linear preferences in money, the result easily extends to more general utility functions $w(t)$. Here, pooling is sustained if the marginal utility from money $w'(t)$ is sufficiently low when evaluated at the wealth constraint $t = t_0$ (in our case normalised to zero).

[23] If the principal refuses to pay the transfer for $H(e_i, b_i, r_i, b_i) > \overline{H}_i$, this may be viewed as an undue restriction of the agent's performance. However, as we show shortly, the effort e_i^* that maximises the agent's utility al-

$$at_i \geq U\left(e_i^*, b_i, r_i, \beta_i\right) - U\left(\bar{e}_i, b_i, r_i, \beta_i\right),\ i = E, I \qquad \text{(MHi)},$$

where $U\left(e_i^*, b_i, r_i, \beta_i\right)$ is the agent's utility when foregoing the transfer and supplying effort $e_i^* = \hat{e}\left(b_i, r_i, \beta_i\right) = \arg\max \hat{U}\left(e_i, b_i, r_i, \beta_i\right)$. The principal chooses $\{\bar{e}_i, b_i, r_i, t_i\}, i = E, I$ so as to maximise $R\big|_{e_i=\bar{e}_i}$, as given in (2), subject to the two moral hazard constraints (MHE) and (MHI). The allocation under full information $\left\{\bar{e}_i^{\,+}, b_i^{\,+}, r_i^{\,+}, t_i^{\,+}\right\} i = I, E$ can then be characterised as follows.

Proposition 4. (i) The principal uses performance pay $t_i^{\,+} = \frac{1}{a}\left[U\left(e^*{}_i, B_i, D_i, \beta_i\right) - U\left(\bar{e}_i, B_i, D_i, \beta_i\right)\right] > 0$ *to stimulate effort beyond the level the agent would choose, i.e.* $\bar{e}_i^{\,+} = ah\left(b_i, r_i, \beta_i\right) + e_i^* > e_i^*$. *(ii) The principal offers a greater reward to the efficient type, i.e.,* $t_E^{\,+} > t_I^{\,+}$, *and this type provides greater effort, i.e.* $\bar{e}_E^{\,+} > \bar{e}_I^{+}$. *(iii) The degree of delegation* D_i^+ *is smaller (greater) than the one implemented with contractible output under full information,* D_i^*, *if* $\alpha > 0$ *(*$\alpha < 0$*).*

Proof: See Appendix.

When output is contractible, the principal can use financial rewards to stimulate effort beyond the level that would be volunteered even by a motivated agent. Notably, as long as agents are intrinsically motivated the rewards need only reflect the agent's utility loss vis-à-vis their utility maximising level of effort, which is significantly less than the full effort cost.[24] The extent to which the principal can stimulate extra effort is governed by the agent's responsiveness to financial rewards, a. As the greater effort (as compared to the situation of non-contractible output) raises the marginal product of the budgetary inputs, the principal also allocates greater total budgets for both types. The degree of delegation is now given by

$$D_i^+ = \frac{\beta_i a + \left(2\beta_i + \alpha\right)u(\cdot)}{\rho a + \left(2\rho - \alpha\right)u(\cdot)} \qquad \text{(14).}^{25}$$

As before, there is over-delegation if and only if the agent has a preference for autonomy, i.e. if and only if $\alpha > 0$, the efficient agent receiving a greater degree of delegation. However, budgets are distorted for a different reason now. In the absence of transfers, the distortion was chosen as to increase the agent's marginal benefit from providing effort. When output is contractible, the principal can stimulate effort more directly by setting an appropriate target $\bar{H}_i$.

---

ways falls short of the effort $\bar{e}_i^{\,*}$ that maximises the principal's surplus. Thus, $H\left(e_i^*, b_i, r_i, \beta_i\right) < \bar{H}_i$ and the target never imposes an upward constraint on the agent's effort. However, as we will see in the next section (see footnote 26), paying the transfer for $H\left(e_E, b_I, r_I, \beta_E\right) > \bar{H}_I$ may in some cases harden the E type's incentive constraint. The principal optimally rules this out by paying the transfer only if the target is met.

[24] This relies on the implicit assumption that financial incentives do not replace the agent's non-financial motivation. In reality, crowding out of intrinsic motivation may matter (Frey 1993, 1997; Bénabou and Tirole 2003, Grepperud and Pedersen 2004). This complication would not alter our more substantive results.

[25] Since $u\left(D_i^{\,+}\right)$ is included on the RHS of (14), this only gives an implicit definition of $D_i^{\,+}$. It is readily verified that a unique value of $D_i^{\,+}$ satisfies the equation.

The distortion in the budgets away from their technologically optimal levels now arises for reasons of rent extraction. By rendering the budget allocation more attractive to the agent, e.g. by granting more autonomy, this helps the principal to contain the transfer payment that is necessary to implement any given effort $\overline{e}_i$. The financial rewards are, thus, leveraged and greater levels of effort can be stimulated. Finally, note that the availability of performance pay allows the principal to reduce the distortion in the degree of delegation away from the technological optimum, where $\left|D_i^* - \tilde{D}_i\right| > \left|D_i^+ - \tilde{D}_i\right|$.

6 Allocation under asymmetric information when output is contractible

Here, the principal can design an output related reward scheme, but she is unable to observe the agents' type. To explore the optimal contract under asymmetric information we introduce some additional notation. Let

$$e_{ij}^* := \hat{e}(b_j, r_j, \boldsymbol{b}_i) = \arg\max U(e_i, b_j, r_j, \boldsymbol{b}_i),\ i, j \in \{E, I\}$$

denote the effort that maximises type i's utility when facing type j's allocation, and let

$$\overline{e}_{ij} := \frac{h(b_j, r_j, \boldsymbol{b}_j)}{h(b_j, r_j, \boldsymbol{b}_i)}\overline{e}_{jj},\ i, j \in \{E, I\}$$

denote the effort type i needs to expend in order to meet type j's performance target $\overline{H}_j = \overline{e}_{jj} h(b_j, r_j, \boldsymbol{b}_j)$. It can then be shown that the agents do not self-select naturally. Specifically, $U(\overline{e}_{II}^+, b_I^+, r_I^+, \boldsymbol{b}_I) + at_I = U(e_{II}^*, b_I^+, r_I^+, \boldsymbol{b})_I < (Ue_{IE}^*, b_E^+, r_E^+, \boldsymbol{b})$, where the inequality follows from $b_E^+ \geq b_I^+$ and $r_E^+ \geq r_I^+$. Hence, the I type could always increase utility by taking the E type's allocation and choosing the preferred effort.[26]

The principal's problem is then to $\max\limits_{\overline{e}_i, b_i, r_i, t_i;\ i=E,I} R$, as given in (2), subject to the moral hazard constraints (MHI) and (MHE), the self-selection constraints

$$U(\overline{e}_{II}, b_I, r_I, \boldsymbol{b}_I) + at_I \geq \max\{U(\overline{e}_{IE}, b_E, r_E, \boldsymbol{b}_I) + at_E, U(e_{IE}^*, b_E, r_E, \boldsymbol{b}_I)\} \qquad \text{(ICI')},$$

$$U(\overline{e}_{EE}, b_E, r_E, \boldsymbol{b}_E) + at_E \geq \max\{U(\overline{e}_{EI}, b_I, r_I, \boldsymbol{b}_E) + at_I, U(e_{EI}^*, b_I, r_I, \boldsymbol{b}_E)\}, \qquad \text{(ICE')},$$

---

[26] It can be shown that

$$\left\{\begin{matrix} U(\overline{e}_{EE}^+, b_E^+, r_E^+, \boldsymbol{b}_E) + at_E \\ = U(e_{EE}^*, b_E^+, r_E^+, \boldsymbol{b}_E) \end{matrix}\right\} > \max\{U(\overline{e}_{EI}, b_I^+, r_I^+, \boldsymbol{b}_E) + at_I, U(e_{EI}^*, b_I^+, r_I^+, \boldsymbol{b}_E)\}$$

implying that the E type self-selects naturally. Incidentally, this illustrates why it is not optimal for the principal to grant performance rewards if output is in excess of the target. It can be checked that $e_{EI}^* > \overline{e}_{EI}$ if a is sufficiently low. In this case the E type would produce $e_{EI}^* h(b_I, r_I, \boldsymbol{b}_E) > \overline{H}_I$ when choosing e_{EI}^*. If the principal rewards output in excess of $\overline{H}_I$, the E type attains utility $U(e_{EI}^*, b_I, r_I, \boldsymbol{b}_E) + at_I > \max\{U(\overline{e}_{EI}, b_I, r_I, \boldsymbol{b}_E) + at_I, U(e_{EI}^*, b_I, r_I, \boldsymbol{b}_E)\}$, implying that the incentive constraint is hardened. It can be shown that $U(e_{EE}^*, b_E^+, r_E^+, b_E) > U(e_{EI}^*, b_I^+, r_I^+, b_{IE}) + at_I$ is not always satisfied.

and the wealth constraints $t_E \geq 0$ and $t_I \geq 0$.

From Proposition 4, part (i), we know that in the case of contractible output it is always optimal to set $t_i > 0; i = E, I$. Hence, we can ignore the wealth constraints. Combining the moral hazard and self-selection constraints, we obtain

$$U\left(\bar{e}_{II}, b_I, r_I, \mathbf{b}_I\right) + at_I \geq \max\left\{\begin{matrix} U\left(\bar{e}_{IE}, b_E, r_E, \mathbf{b}_I\right) + at_E, \\ U\left(e^*_{IE}, b_E, r_E, \mathbf{b}_I\right); U\left(e^*_{II}, b_I, r_I, \mathbf{b}_I\right) \end{matrix}\right\} \quad (15a),$$

$$U\left(\bar{e}_{EE}, b_E, r_E, \mathbf{b}_E\right) + at_E \geq \max\left\{\begin{matrix} U\left(\bar{e}_{EI}, b_I, r_I, \mathbf{b}_E\right) + at_I, \\ U\left(e^*_{EI}, b_I, r_I, \mathbf{b}_E\right), U\left(e^*_{EE}, b_E, r_E, \mathbf{b}_E\right) \end{matrix}\right\} \quad (15b).$$

In order to identify the relevant constraints let us, for a moment, consider all six candidate constraints, where m^i_1, m^i_2 and m^i_3, $i = E, I$ denote the shadow prices for the constraint relating to the first, second and third element in bracelets on the RHS of (15a) and (15b), respectively. The first-order conditions with respect to $\{t_i, \bar{e}_{ii}, b_i, r_i\}$, $i = E, I$ are then given by

$$-(1 - l) + a\left(m^I_1 + m^I_2 + m^I_3 - m^E_1\right) = 0 \quad (16a), \qquad -l + a\left(m^E_1 + m^E_2 + m^E_3 - m^I_1\right) = 0 \quad (16b),$$

$$(1 - l)\left[h\left(b_I, r_I, \mathbf{b}_I\right) + \tfrac{1}{a}U_e\left(\bar{e}_{II}, b_I, r_I, \mathbf{b}_I\right)\right] - m^E_1\left[U_e\left(\bar{e}_{EI}, b_I, r_I, \mathbf{b}_E\right) - U_e\left(\bar{e}_{II}, b_I, r_I, \mathbf{b}_I\right)\right] = 0 \quad (16c),$$

$$l\left[h\left(b_E, r_E, \mathbf{b}_E\right) + \tfrac{1}{a}U_e\left(\bar{e}_{EE}, b_E, r_E, \mathbf{b}_E\right)\right] - m^I_1\left[U_e\left(\bar{e}_{IE}, b_E, r_E, \mathbf{b}_I\right) - U_e\left(\bar{e}_{EE}, b_E, r_E, \mathbf{b}_E\right)\right] = 0 \quad (16d),$$

$$\left\{\begin{matrix} (1 - l)\left[\bar{e}_{II} h_x\left(b_I, r_I, \mathbf{b}_I\right) + \tfrac{1}{a}U_x\left(\bar{e}_{II}, b_I, r_I, \mathbf{b}_I\right) - 1\right] - m^I_3 U_x\left(e^*_{II}, b_I, r_I, \mathbf{b}_I\right) \\ - m^E_1\left[U_x\left(\bar{e}_{EI}, b_I, r_I, \mathbf{b}_E\right) - U_x\left(\bar{e}_{II}, b_I, r_I, \mathbf{b}_I\right)\right] - m^E_2 U_x\left(e^*_{EI}, b_I, r_I, \mathbf{b}_E\right) \end{matrix}\right\} = 0;\ x = b, r \quad (16e/f),$$

$$\left\{\begin{matrix} l\left[\bar{e}_{EE} h_x\left(b_E, r_E, \mathbf{b}_E\right) + \tfrac{1}{a}U_x\left(\bar{e}_{EE}, b_E, r_E, \mathbf{b}_E\right) - 1\right] - m^E_3 U_x\left(e^*_{EE}, b_E, r_E, \mathbf{b}_E\right) \\ - m^I_1\left[U_x\left(\bar{e}_{IE}, b_E, r_E, \mathbf{b}_I\right) - U_x\left(\bar{e}_{EE}, b_E, r_E, \mathbf{b}_E\right)\right] - m^I_2 U_x\left(e^*_{IE}, b_E, r_E, \mathbf{b}_I\right) \end{matrix}\right\} = 0;\ x = b, r \quad (16g/h).$$

As is common for models of mixed adverse selection and moral hazard, there is no clear-cut rule to determine the binding constraints within (15a) and (15b) and, thus, to determine which of the shadow prices should be positive.

Case 1: $m^E_1 = m^I_1 = 0$

If $m^E_1 = m^I_1 = 0$ neither type has an incentive to mimic the other in order to attain the performance reward. We can then characterise the allocation under asymmetric information, $\left\{t_i^{++}, \bar{e}_{ii}^{++}, B_i^{++}, D_i^{++}\right\}$, $i = E, I$, as follows.

Proposition 5. (i) There is no distortion in the performance targets/effort levels, i.e. $\bar{e}^{++}_i = ah\left(b_i, r_i, \mathbf{b}_i\right) + e^*_i$. *The budget allocation follows a pattern similar to figure 3. Specifically, (ii) the principal chooses the transfers* $t_i^{++} = \tfrac{1}{a}\left[U\left(e^*{}_i, B_i, D_i, \mathbf{b}_i\right) - U\left(\bar{e}_i, B_i, D_i, \mathbf{b}_i\right)\right]$ *to* *address moral hazard alone for* $a \in \left[0, \bar{a}_i^+(\alpha)\right]$, where $\bar{a}_i^+(\alpha) \in (0, \infty)$; $i = E, I$, *(iii) and budgets are pooled, i.e.* $B_I^{++} = B_E^{++} = B_P$ *and* $D_I^{++} = D_E^{++} = D_P$, *for all pairs* (a, α) *satisfying* $a \in \left[0, \bar{a}_i^+(\alpha)\right]$, $i = E, I$, and $\alpha > \hat{\alpha}$, *with* $\hat{\alpha}$ *as defined in (12).*[27]

Proof: See Appendix.

Performance targets (conditional on the budget allocation) remain undistorted. This is not surprising as in the present case we have assumed that a misrepresentation of type is not motivated by a desire to attain a more preferred performance target/payment. For the same reason the budget allocation is similar to the one realised in the case of non-contractible output (see figure 3 and Proposition 3). If the preference for financial rewards is weak, the principal uses them only to induce effort but not to facilitate the agents' self-selection. In this case the budgets are pooled if the preference for autonomy is sufficiently high (but not too high). As discussed before, pooling arises when rent-extraction conflicts with the E type's self-selection constraint, while the cost of financial transfers exceeds the gains from a more efficient separating allocation. We conclude by establishing a condition for the allocation just described.

Lemma 8: The allocation $\left\{t_i^{++}, \bar{e}_{ii}^{++}, B_i^{++}, D_i^{++}\right\}$, $i = E, I$ *is implemented if and only if* $\bar{e}_{II} \le e_{IE}^* \sqrt{v}$, *with* $v := \frac{U(e_{EE}^*, b_E, r_E, \beta_E) - U(e_{IE}^*, b_E, r_E, \beta_I)}{U(e_{EI}^*, b_I, r_I, \beta_E) - U(e_{II}^*, b_I, r_I, \beta_I)} \ge 1$.

Proof: See Appendix

This requires that the effort level assigned to the I type be sufficiently low. Otherwise, the E type has an incentive to mimic the I type and capture the performance payment at a low effort. Since $\bar{e}_{II}^{++} = ah(b_I, r_I, \beta_I) + e_{II}^*$ the condition is satisfied if (and only if) the preference for autonomy, a, is not too large.

Case 2: $m_1^E > m_1^I = 0$

Suppose now $\bar{e}_{II} > e_{IE}^* \sqrt{v}$ so that $m_1^E > 0$, implying that the E type has an incentive to mimic the I type. The following can then be shown.

Lemma 9: $m_1^I = 0$ *if and only if* $\left\{\left[\frac{h(b_E, r_E, \beta_E)}{h(b_E, r_E, \beta_I)}\right]^2 - 1\right\}\bar{e}_{EE}^2 \ge \left\{1 - \left[\frac{h(b_I, r_I, \beta_I)}{h(b_I, r_I, \beta_E)}\right]^2\right\}\bar{e}_{II}^2$.

Proof: See Appendix.

In this case, the I type does not aspire to attain the E type's performance reward. Transfers are then given by $t_E = \frac{1}{\alpha}\left[U(\bar{e}_{EI}, b_I, r_I, \beta_E) - U(\bar{e}_{EE}, b_E, r_E, \beta_E)\right]$ and $t_I = \frac{1}{\alpha}\left[U(e_{IE}^*, b_E, r_E, \beta_I) - U(\bar{e}_{II}, b_I, r_I, \beta_I)\right]$. It is easy to establish from (16a)-(16h) the following features of the equilibrium allocation.

[27] Note that $a \in \left[0, \bar{a}_i^+(\alpha)\right]$, $i = E, I$, and $\alpha > \hat{\alpha}$ are sufficient for pooling. Pooling also arises for some $\alpha < \hat{\alpha}$. Thus, the presence of performance pay makes pooling more likely. This is due to the relative increase in the E type's budgets in the presence of performance pay.

Proposition 6. Consider $m_1^E > m_1^I = 0$. *(i) Then, the E type's effort* $\bar{e}_E^{+++}$ *is not distorted away from the full information level but the I type's effort* $\bar{e}_I^{+++}$ *is distorted downwards. (ii) IC constraints bind for both types. (iii) The E type's budget allocation is as depicted in Proposition 5, and (iv) the I type's total budget* B_I^{+++} *and the degree of delegation* D_I^{+++} *are distorted downwards from the levels described in Proposition 5.*

Proof: See Appendix.

If the preference for financial rewards is sufficiently high, the E type has an incentive to mimic the I type and reap this type's performance reward. This conforms to the set-up of many standard models with 'false moral hazard' and adverse selection (e.g. Laffont and Martimort 2002: section 7.4.1). Here, the principal's response is to lower the I type's target (and effort) as well as the degree of delegation assigned to this type. In so doing, she makes the I type's allocation less attractive to the E type and extracts some of the informational rent. In contrast to the standard model, however, the E type's allocation remains distorted as described in Proposition 5. This is because the self-selection constraint for the I type remains binding. Indeed, in this particular instance adverse selection incentives arise for both types. While the E type seeks to obtain the more attractive package of performance pay offered to the I type, the I type seeks to obtain the more attractive budget allocation designated to the E type without aiming to attain the performance target. The principal responds to the compounded incentive problem by using the I type's performance target as an additional instrument.

7 Heterogeneity in the preference for autonomy

We have carried out our analysis assuming that agent heterogeneity mainly relates to the productivity of the agent or the project they are carrying out. At the same time we have maintained that agents have similar preferences for autonomy. It has been suggested that agents self-select into organisations (or occupations) according to their preference for job characteristics rather than their productivity (Dixit 2002 Besley and Ghatak 2003). In this case, productivity becomes the distinguishing characteristic within an organisation as we have assumed. Alternatively, one could argue that agents working within the same organisation substantively differ in the degree to which they prefer autonomy. We will now demonstrate that this case does not lead to significantly different results. Assume thus that agents differ in their preference for autonomy, where $a \in \{a_L, a_H\}$, with $a_L < a_H$. It is then readily verified that $\hat{D}_i = \frac{b+a_i}{r-a_i} = \arg\max \hat{U}(\cdot, a_i)$ and $D_i^* = \frac{2b+a_i}{2r-a_i} = \arg\max \hat{H}(\cdot, a_i)$, where $i = H, L$. As before over delegation vis-à-vis the technological optimum $\tilde{D} = \frac{b}{r}$ occurs under complete information if and only if agents have a preference for autonomy, i.e. $D_i^* > \tilde{D} \Leftrightarrow a_i > 0$. At the same time, the 'better motivated' H type receives greater budgets $b_H \geq b_L$, $r_H \geq r_L$ along both the controlled and delegated dimension, causing adverse selection incentives on the part of the L type. Supposing that financial transfers are not used (small a), the allocation $\{D_i^{**}, B_i^{**}\}$ under asymmetric information can be summarised as follows.

Proposition 7. (i) There is over-delegation to the L type, $D_L^{**} > D_L^*$, *if and only if* $a_L > 0$. *(ii) There is under-delegation to the H type,* $D_H^{**} < D_H^*$, *if and only if* $a_L > \frac{1}{2} a_H$. *(iii) There exists* $\underline{a} \in (\frac{1}{2} a_H, a_H]$ *such that pooling ensues for* $a_L \in [\underline{a}, a_H]$ *if* $a_H \ln D_H^* \geq 1$.

Proof: See Appendix.

These findings are notable on two grounds. For $a_L \geq 0$, the distortion in the allocation decreases with the degree of heterogeneity between agents, as measured by the spread $a_H - a_L$. Second, pooling may then become attractive if the agents become sufficiently similar. This is the case when the preference for autonomy on the part of the H type a_H is sufficiently strong and/or if delegation is productive such that $b >> r$, implying a high D_H^*. Hence, pooling is the outcome if delegation is strongly preferred by all types as well as by the principal.

8 Conclusions

We have studied a principal agent model of budget delegation to illuminate some of the implications of agents' preferences for autonomy. The principal faces both a moral hazard and adverse selection problem to which she can react by adjusting the levels of budgets allocated to an agent, the control she retains as well as a payment to the agent. Under full information about agent's productivity the principal adjusts the budget and its delegation in order to stimulate additional effort or – in case of performance pay – in order to reduce the payment. Our key results relate to the case in which the principal is uninformed about the agent's type, the productivity under delegation. Contracts turn out to be sensitive to the agent's preference for autonomy and to the trade-off between job satisfaction and financial transfers. Generally, if agents strongly prefer job satisfaction to financial transfers, the latter will not be used in order to facilitate self-selection. A separating allocation can then be attained only if preferences for autonomy are not too strong. The distortion in the degree of delegation depends on the agents' preference for autonomy. If agents are indifferent to autonomy, a separating allocation requires over-delegation to the efficient type and no distortion for the inefficient type. For an increasing preference for autonomy the gap in the degree of delegation is gradually closed and a pooling allocation is reached when the preference for autonomy is sufficiently strong. This form of non-responsiveness arises when the agent's and the principal's preferences regarding the allocation of autonomy are sufficiently divergent. Separation remains feasible when the agent's preference for financial transfers is sufficiently strong. In this case, strong preferences for autonomy lead to a reversal in the degree of delegation. These results highlight the important implications of divergent preferences with regard to the mode of production.

From an empirical perspective one would expect that budget allocations will be less sensitive to differences in the productivity of agents managing a delegated budget within those organisations or organisational units that rely on agents with a strong preference for autonomy. Indeed, pooling is the likely outcome if strong preferences for autonomy are coupled with weak interest in financial rewards relative to job-satisfaction. Cursory evidence suggests that this ties-in rather well with the prevalence of weaker performance incentives within such organisations. (Roomkin and Weisbrod 1999, Francois 2000, Dixit 2002, Besley and Ghatak 2003). A tendency towards pooled budgets within such organisations may be viewed critically in that it implies misallocation of funds both to productive and unproductive agents. However, it should be borne in mind that pooling is an optimal response to the agents' strong focus on job-satisfaction, which could only be swung by excessive financial transfers. Furthermore, the productivity gains from intrinsically motivated agents may well over-compensate the lack of efficiency in the budget allocation.

While we derive the main results for the case in which output (and effort) are non-contractible, we show that their substance carries over to the case of contractible output. We also show that similar results hold in a situation in which agent heterogeneity pertains with regard to the preference for autonomy. A number of limitations and possible extensions deserve discussion. First, we assume that the agent's input, effort, and the principal's inputs, the budgets, are complements in production. This implies that the principal, acting as first-mover, can stimulate additional effort by increasing the budgets. More generally, budgets and effort may also be substitutes, in case of which the principal would reduce her own input.[28] However, as long as the efficient type receives the greater budget and greater degree of delegation, this would not substantively alter our analysis of budgeting under asymmetric information. Second, we assume that productivity is the agent's private information and the principal uses the budgets to screen agents. Alternatively, one could follow Bénabou and Tirole (2003) and assume that the principal is informed about the productivity of the project and thus its value to the agent. In this case, the principal could signal to the agent a productive/attractive project by over-budgeting and over-delegating to an extent that would not be profitable for the non-productive project. However, separation may not be possible if a high preference for autonomy on the part of the agent and the ensuing effort incentives make it profitable to over-delegate to all types. Third, we have disregarded the effects of uncertainty. One way of introducing risk into the present framework is to model output as a random variable, where the agent's type b corresponds to the probability of making effective use of the delegated budget. As risk is attached to the use of the delegated budget more than to the use of the regulated budget the principal has an incentive to under-delegate relative to the first-best in order to reduce the agent's risk premium. The presence of risk-aversion may have a positive bearing on the feasibility of separation. As the inefficient agent faces a greater probability of failure under delegation, aversion to this risk reduces the incentive to aspire for the efficient agent's allocation. Thus, for high levels of a risk contributes to reducing the distortion in the separating allocation. If agents are very risk averse, the presence of risk may even reverse the incentive problem.

9 References

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

Andreoni, J. (1990), 'Impure Altruism and Donations to Public Goods: A Theory of Warm-glow Giving', Economic Journal 100, 464-477.

Barkema, H.G. (1995), 'Do managers work harder when they are monitored?', Kyklos 48, 19-42.

Bénabou, R. and J. Tirole (2003), 'Intrinsic and Extrinsic Motivation', Review of Economic Studies 70, 489-520.

Bernardo, A.E., H. Cai and J. Luo (2001), 'Capital budgeting and compensation with asymmetric information and moral hazard', Journal of Financial Economics 61, 311-344.

Besley, T. and M. Ghatak (2003), Competition and Incentives with Motivated Agents, mimeo.

Caillaud, B., B. Jullien and P. Picard (1996), 'Hierarchical organization and incentives', European Economic Review 40, 687-695.

[28] See Glazer (2004) for an analysis of both cases.

Delfgaauw, J. and R.A.J. Dur (2002), Signalling an Screening of Workers' Motivation, University of Rotterdam: Tinbergen Institute Discussion Paper TI 2002-050/3.

Dewatripont, M., I. Jewitt and J. Tirole (1999), 'The Economics of Career Concerns, Parts I and II', Review of Economic Studies 66, 183-217.

Dixit, A. (2002), 'Incentives and Organizations in the Public Sector: An Interpretative Review', Journal of Human Resources 37, 696-727.

Encinosa, III, W.E., M. Gaynor and J.B. Rebitzer (1997), 'The Sociology of Groups and the Economics of Incentives: Theory and Evidence on Compensation Systems', NBER Working Paper 5953.

Francois, P. (2000), '"Public Service Motivation" as an argument for government provision', Journal of Public Economics 78, 275-299.

Frey, B.S. (1993), 'Shirking and Work Morale? The Impact of Regulating', European Economic Review 37, 1523-1532.

Frey, B.S. (1997), 'On the relationship between intrinsic and extrinsic work motivation', International Journal of Industrial Organization 15, 427-439.

Frey, B.S. and M. Benz (2002), Being Independent is a Great Thing: Subjective Evaluations of Self-Employment and Hierarchy, University of Zurich: Working Paper No. 135.

Gertner, R.H., D.S. Scharfstein and J.C. Stein (1994), 'Internal versus External Capital Markets', Quarterly Journal of Economics 109, 1211-1230.

Glazer, A. (2004), 'Motivating Devoted Workers', International Journal of Industrial Organization 22, 427-440.

Grepperud, S. and P.A. Pedersen (2004), 'Crowding Effects and Work Ethics', forthcoming in Labor.

Guesnerie, R. and J.-J. Laffont (1984), 'A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm', Journal of Public Economics 25, 329-369.

Harris, M. and A. Raviv (1998), 'Capital budgeting and delegation', Journal of Financial Economics 50, 259-289.

Heckman, J.A., J.A. Smith and C. Taber (1996), What do Bureaucrats do? The Effects of Performance Standards and Bureaucratic Preferences on Acceptance into the JTPA Programme, Cambridge, MA: NBER Working Paper 5535.

Holmstrom, B. (1999), 'Managerial Incentive Problems: A Dynamic Perspective', Review of Economic Studies 66, 169-182.

Holmstrom, B. and P. Milgrom (1994), 'The Firm as an Incentive System', American Economic Review 84, 972-991.

Laffont, J.-J. and D. Martimort (2002), The Theory of Incentives: the Principal-Agent Model' Princeton and Oxford: Princeton UP.

McGuire, T.P. (2000), 'Physician Agency', in: A.J. Culyer and J.P. Newhouse, Handbook of Health Economics (Vol. 1), Amsterdam: North-Holland.

Melumad, N., D. Mookherjee and S. Reichelstein (1997), 'Contract Complexity, Incentives, and the Value of Delegation', Journal of Economics and Management Strategy 6, 257-289.

Mitusch, K. (2000), 'Job Independence as an Incentive Device' Economica 67, 245-263.

Morand, P.-H. and L. Thomas (2003), 'On Non-responsiveness in Adverse Selection Models with Common Value', Topics in Theoretical Economics 3(1), Article 3.

Murdock, K. (2002), Intrinsic Motivation and Optimal Incentive Contracts, RAND Journal of Economics 33, 650-671.

Roomkin, M.J. and B.A. Weisbrod (1999), 'Managerial Compensation and Incentives in For-Profit and Non-profit Hospitals', Journal of Law, Economics and Organization 15, 750-81.

Sen, A. (1997), 'Maximization and the Act of Choice', Econometrica 65, 745-779.

10 Appendix

Proof of Proposition 1:

(i) Follows from comparative static analysis for the system (7a) and (7b) under use of (4).

(ii) and (iii) Using (7a) and (7b) together with the appropriate derivatives from (4) we obtain $\frac{\hat{H}_r^i(b,r)}{\hat{H}_b^i(b,r)} = \frac{(2r-a)b_i^*}{(a+2b_i)r_i^*} = 1 \Leftrightarrow D_i^* = \frac{b_i^*}{r_i^*} = \frac{a+2b_i}{2r-a};\quad i = E, I$. Obviously then, $\frac{dD_i^*}{da} > 0$ and $\frac{dD_i^*}{db_i} > 0$ implying the statements in part (ii) and (iii).■

Proof of Lemma 1:

(i) We prove $m_1 > 0$ by contradiction. Suppose $m_1 = 0$. For $m_2 \geq 0$, it follows from (9c)-(9f) that the equilibrium values $\{b_E^{**}, r_E^{**}, b_I^{**}, r_I^{**}\}$ satisfy $b_E^{**} \geq b_E^*; r_E^{**} \geq r_E^*; b_I^{**} \leq b_I^*; r_I^{**} \leq r_I^*$, where the values with the single asterisk correspond to the unconstrained problem. Hence, $\hat{U}(b_E^{**}, r_E^{**}, \boldsymbol{b}_I) \geq \hat{U}(b_E^*, r_E^*, \boldsymbol{b}_I) > \hat{U}(b_I^*, r_I^*, \boldsymbol{b}_I) \geq \hat{U}(b_I^{**}, r_I^{**}, \boldsymbol{b}_I)$, where the strict inequality has been shown to hold true for the unconstrained solution. But then, $\hat{U}(b_E^{**}, r_E^{**}, \boldsymbol{b}_I) > \hat{U}(b_I^{**}, r_I^{**}, \boldsymbol{b}_I)$, which contradicts $m_1 = 0$.

To prove $m_1 > m_2$ by contradiction suppose $m_2 \geq m_1$. From (9a) this implies $m_3 > 0$ and, therefore, $t_I = 0$. (ICE) and (ICI) then imply

$$at_E = \hat{U}(b_I^{**}, r_I^{**}, \boldsymbol{b}_E) - \hat{U}(b_E^{**}, r_E^{**}, \boldsymbol{b}_E) = \hat{U}(b_I^{**}, r_I^{**}, \boldsymbol{b}_I) - \hat{U}(b_E^{**}, r_E^{**}, \boldsymbol{b}_I) \geq 0 \qquad \text{(A1)},$$

where the inequality follows from the wealth constraint. Using (9c) and (9d), we obtain

$$0 = (1-l)\left[\hat{H}_b(b_I, r_I, \boldsymbol{b}_I) - 1\right] + m_1\hat{U}_b(b_I, r_I, \boldsymbol{b}_I) - m_2\hat{U}_b(b_I, r_I, \boldsymbol{b}_E)$$

$$\leq (1-l)\left[\hat{H}_b(b_I, r_I, \boldsymbol{b}_I) - 1\right] + m_1\overbrace{\left[\hat{U}_b(b_I, r_I, \boldsymbol{b}_I) - \hat{U}_b(b_I, r_I, \boldsymbol{b}_E)\right]}^{<0} < (1-l)\left[\hat{H}_b(b_I, r_I, \boldsymbol{b}_I) - 1\right],$$

and, likewise,

$$0 = (1-l)\left[\hat{H}_r(b_I, r_I, \boldsymbol{b}_I) - 1\right] + m_1\hat{U}_r(b_I, r_I, \boldsymbol{b}_I) - m_2\hat{U}_r(b_I, r_I, \boldsymbol{b}_E) < (1-l)\left[\hat{H}_r(b_I, r_I, \boldsymbol{b}_I) - 1\right]$$

implying $b_I^{**} < b_I^*$ and $r_I^{**} < r_I^*$. Similarly, one obtains from (9e) and (9f) $b_E^{**} > b_E^*$ and $r_E^{**} > r_E^*$. Together with $\hat{U}(b_E^*, r_E^*, \boldsymbol{b}_i) > \hat{U}(b_I^*, r_I^*, \boldsymbol{b}_i)$ this implies $\hat{U}(b_I^{**}, r_I^{**}, \boldsymbol{b}_i) - \hat{U}(b_E^{**}, r_E^{**}, \boldsymbol{b}_i) < 0$, which contradicts (A1).

(ii) $m_4 > 0$ follows from (9b) under observation of $m_1 > m_2$.

(iii) Assuming $D_E > D_I$ we use figure 2a to show that this contradicts $m_2 > 0$.[29] Consider the allocations $E = \{B_E, D_E\}$ and $I = \{B_I, D_I\}$, obviously satisfying $D_E > D_I$. Since $m_1 > m_2$, it follows from $m_2 > 0$ that both (ICE) and (ICI) bind simultaneously. Noting $m_4 > 0 \Leftrightarrow t_E = 0$ and the constraint $t_I \geq 0$ it must be true that $at_I = \hat{U}(E, b_I) - \hat{U}(I, b_I) = \hat{U}(E, b_E) - \hat{U}(I, b_E) \geq 0$. By construction, the allocations $E' = \{B_E', D_I\}$ and $E'' = \{B_E'', D_I\}$ satisfy $\hat{U}(E, b_I) = \hat{U}(E', b_I)$ and $\hat{U}(E, b_E) = \hat{U}(E'', b_E)$. Consequently,

$$at_I = \hat{U}(E', b_I) - \hat{U}(I, b_I) = \hat{U}(E'', b_E) - \hat{U}(I, b_E) \geq 0 \qquad \text{(A2)}.$$

For not too large differences in the budgets $B_E'' - B_I > 0$ and $B_E' - B_I > 0$, the second equality in (A2) implies $\hat{U}_B(\cdot, b_E)(B_E'' - B_I) = \hat{U}_B(\cdot, b_I)(B_E' - B_I)$. Since $\hat{U}_B(\cdot, b_E) > \hat{U}_B(\cdot, b_I)$ this implies $B_E' > B_E''$, a contradiction. Hence, $m_2 > 0$ is incompatible with $D_E > D_I$.

We go on to show that $m_2 > 0$ contradicts $D_E = D_I$ and $B_E \neq B_I$. To see this, consider the allocations E and $I' = \{B_I', D_E\}$ in figure 2a. Here, $m_2 > 0$ implies $at_I = \hat{U}(E, b_I) - \hat{U}(I', b_I) = \hat{U}(E, b_E) - \hat{U}(I', b_E) \geq 0$. For small differences in budgets $B_E - B_I' > 0$, this implies $\hat{U}_B(\cdot, b_E)(B_E - B_I') = \hat{U}_B(\cdot, b_I)(B_E - B_I')$. For $\hat{U}_B(\cdot, b_E) > \hat{U}_B(\cdot, b_I)$ this is satisfied if and only if $B_E = B_I'$, a contradiction.

(iv) Assume by contradiction that $m_3 > 0$ and $D_E < D_I$ are simultaneously true. Noting that $m_3 > 0$ implies $t_I = 0$, it follows that (ICI) requires $\hat{U}(B_I, D_I, b_I) - \hat{U}(B_E, D_E, b_I) = 0$. This would, indeed, be satisfied by the allocations $E = \{B_E, D_E\}$ and $E' = \{B_I, D_I\}$ in figure 2b. However, it is immediately established that $\hat{U}(E, b_I) - \hat{U}(E', b_I) = 0 \Rightarrow \hat{U}(E, b_E) - \hat{U}(E', b_E) < 0$ implying that (ICE) is violated. Hence, $m_3 > 0$ and $D_E < D_I$ cannot hold at the same time. Consider now $D_E = D_I$. Since $m_3 > 0$ implies $t_I = t_E = 0$, it follows from (ICI) and (ICE) that $B_E = B_I$. Hence, $D_E = D_I$ and $B_E \neq B_I$ imply $m_3 = 0$.■

Proof of Lemma 2: Equating the LHS of (9c) and (9d), and rearranging gives

$$\left\{ \begin{array}{l} \hat{H}_r(b_I, r_I, b_I)\left[\frac{\hat{H}_b(b_I, r_I, b_I)}{\hat{H}_r(b_I, r_I, b_I)} - 1\right] + \frac{m_1}{1-l}\hat{U}_r(b_I, r_I, b_I)\left[\frac{\hat{U}_b(b_I, r_I, b_I)}{\hat{U}_r(b_I, r_I, b_I)} - 1\right] \\ + \frac{m_2}{1-l}\hat{U}_r(b_I, r_I, b_E)\left[1 - \frac{\hat{U}_b(b_I, r_I, b_E)}{\hat{U}_r(b_I, r_I, b_E)}\right] \end{array} \right\} = 0.$$

Adding $1 - \frac{\hat{H}_b(b_I, r_I, b_I)}{\hat{H}_r(b_I, r_I, b_I)}$ to both sides yields

[29] Taking the appropriate derivatives from (5) one obtains

$$\left.\frac{db}{dr}\right|_{\hat{U}(b_E)=\overline{U}} = \frac{-(r-a)b}{(a+b_E)r} > \frac{-(r-a)b}{(a+b_I)r} = \left.\frac{db}{dr}\right|_{\hat{U}(b_I)=\overline{U}},$$

implying that for every pair (r, b) the I type's indifference curves are steeper sloped than the E type's.

$$1-\frac{\hat{H}_b(b_I,r_I,b_I)}{\hat{H}_r(b_I,r_I,b_I)}=\left\{\begin{matrix}\left[1-\hat{H}_r(b_I,r_I,b_I)\right]\left[1-\frac{\hat{H}_b(b_I,r_I,b_I)}{\hat{H}_r(b_I,r_I,b_I)}\right]+\frac{m_1}{1-l}\hat{U}_r(b_I,r_I,b_I)\left[\frac{\hat{U}_b(b_I,r_I,b_I)}{\hat{U}_r(b_I,r_I,b_I)}-1\right]\\+\frac{m_2}{1-l}\hat{U}_r(b_I,r_I,b_E)\left[1-\frac{\hat{U}_b(b_I,r_I,b_E)}{\hat{U}_r(b_I,r_I,b_E)}\right]\end{matrix}\right\}.$$

$$=\left\{\begin{matrix}\frac{m_1}{1-l}\hat{U}_r(b_I,r_I,b_I)\left[\frac{\hat{U}_b(b_I,r_I,b_I)}{\hat{U}_r(b_I,r_I,b_I)}-\frac{\hat{H}_b(b_I,r_I,b_I)}{\hat{H}_r(b_I,r_I,b_I)}\right]\\+\frac{m_2}{1-l}\hat{U}_r(b_I,r_I,b_E)\left[\frac{\hat{H}_b(b_I,r_I,b_I)}{\hat{H}_r(b_I,r_I,b_I)}-\frac{\hat{U}_b(b_I,r_I,b_E)}{\hat{U}_r(b_I,r_I,b_E)}\right]\end{matrix}\right\}. \tag{A3}$$

where the second equality follows under observation of $1-\hat{H}_r(b_I,r_I,b_I)=\frac{m_1}{1-l}\hat{U}_r(b_I,r_I,b_I)-\frac{m_2}{1-l}\hat{U}_r(b_I,r_I,b_E)$. as from (9d). Inserting the relevant derivatives from (4) and (5) into (A3) and rearranging yields

$$D_I^{**}=\frac{b_I^{**}}{r_I^{**}}=\frac{2b_I+a}{2r-a}+\frac{m_1}{1-l}\hat{U}_r(b_I,r_I,b_I)\left(\frac{b_I+a}{r-a}-\frac{2b_I+a}{2r-a}\right)+\frac{m_2}{1-l}\hat{U}_r(b_I,r_I,b_E)\left(\frac{2b_I+a}{2r-a}-\frac{b_E+a}{r-a}\right).$$

Employing the definitions in (6) and (8) then gives the term in (10a). Starting from (9e) and (9f), it is straightforward to derive (10b) in a similar fashion.■

Proof of Lemma 3:

(i) From (6) and (8), $a<0\Leftrightarrow\hat{D}_I<D_I^*$. For $m_2=0$ it follows from (10a) and (10b) that $D_I^{**}<D_I^*$ and $D_E^*<D_E^{**}$. Since $D_E^*<D_I^*$, it follows immediately that $D_I^{**}<D_I^*<D_E^*<D_E^{**}$, where the inequalities imply $m_2=0$ from part (iii) Lemma 1.

(ii) Since $a\geq 0\Leftrightarrow D_I^*\leq\hat{D}_I$, it follows from (10a) that $D_I^*\leq D_I^{**}$ for $m_2=0$. From (6), (8) and (11), we obtain that $a<\overline{a}\Leftrightarrow\hat{D}_I<D_E^*$. For $m_2=0$, it then follows from (10b) that $D_E^*<D_E^{**}$. Hence, $a\in[0,\overline{a}[\Leftrightarrow D_I^*\leq D_I^{**}<D_E^*<D_E^{**}$, where the middle inequality is consistent with $m_2=0$.

(iii) and (iv) Since $a\geq\overline{a}\Leftrightarrow D_E^*\leq\hat{D}_I$, it follows from (10b) that $D_E^{**}\leq D_E^*$ for $m_2=0$. With $D_I^*<D_I^{**}$, we have to prove that $a<\hat{a}\Leftrightarrow D_E^{**}>D_I^{**}$. Using (10a) and (10b) we obtain

$$D_E^{**}\geq D_I^{**}\Leftrightarrow\left\{\begin{matrix}\left[1+\frac{m_1\hat{U}_r(b_E,r_E,b_I)}{l}-\frac{m_2\hat{U}_r(b_E,r_E,b_E)}{l}\right]D_E^*-\left[1-\frac{m_1\hat{U}_r(b_I,r_I,b_I)}{1-l}+\frac{m_2\hat{U}_r(b_I,r_I,b_E)}{1-l}\right]D_I^*\\-m_1\left[\frac{\hat{U}_r(b_E,r_E,b_I)}{l}+\frac{\hat{U}_r(b_I,r_I,b_I)}{1-l}\right]\hat{D}_I+m_2\left[\frac{\hat{U}_r(b_E,r_E,b_E)}{l}+\frac{\hat{U}_r(b_I,r_I,b_E)}{1-l}\right]\hat{D}_E\end{matrix}\right\}\geq 0,$$

or after employing (9d) and (9f) and rearranging

$$D_E^{**}\geq D_I^{**}\Leftrightarrow Z+m_2\left[\frac{\hat{U}_r(b_E,r_E,b_E)}{l}+\frac{\hat{U}_r(b_I,r_I,b_E)}{1-l}\right]\left(\hat{D}_E-\hat{D}_I\right)\geq 0 \tag{A4}$$

$$Z=\hat{H}_r(b_E,r_E,b_E)\left(D_E^*-\hat{D}_I\right)-\hat{H}_r(b_I,r_I,b_I)\left(D_I^*-\hat{D}_I\right).$$

Inserting the appropriate derivatives from (4) and rearranging terms we can rewrite

$$Z = \tfrac{k}{r-a}\begin{Bmatrix} (r-a)\left[(2b_E+a)b_E^{2b_E+a}r_E^{2r-a-1}-(2b_I+a)b_I^{2b_I+a}r_I^{2r-a-1}\right] \\ -\left(b_E^{2b_E+a}r_E^{2r-a-1}-b_I^{2b_I+a}r_I^{2r-a-1}\right)(2r-a)(b_I+a) \end{Bmatrix}.$$

Assume for the moment $m_2 = 0$ and consider a pooling allocation $\{b_E = b_I = b_P; r_E = r_I = r_P\}$ that satisfies

$$Z\Big|_{\{b_E=b_I=b_P\,;\, r_E=r_I=r_P\}} = \tfrac{kr_P^{2r-a-1}b_P^a}{r-a}\begin{Bmatrix} (r-a)\left[(2b_E+a)b_P^{2b_E}-(2b_I+a)b_P^{2b_I}\right] \\ -\left(b_P^{2b_E}-b_P^{2b_I}\right)(2r-a)(b_I+a) \end{Bmatrix} = 0 .$$

It is readily verified that this implies $a = \frac{2(b_E-b_I)rb_P{}^{2b_E}}{(r+2b_E-b_I)b_P{}^{2b_E}-(r+b_I)b_P{}^{2b_I}} =: \hat{a}$. Note that $\hat{a} \in \left]\overline{a}, r\right[$. From comparative static analysis of the system $\hat{U}(b_I, r_I, b_I) - \hat{U}(b_E, r_E, b_I) = 0$, where $t_I = 0$, (9a) and (9c)-(9f) each with $m_2 = 0$ we obtain after tedious calculations $\left(\frac{dD_E^{**}}{da} - \frac{dD_I^{**}}{da}\right)\Big|_{D_E=D_I=D_P;B_E=B_I=B_P;m_2=0} < 0$. Note that for $a = 0$ and, therefore, for $t_E = t_I = 0$, (ICI) and (ICE) imply $D_E = D_I \Leftrightarrow B_E = B_I$. Hence, $\frac{dZ}{da}\Big|_{a=0;Z=0;m_2=0} < 0$. But then,

$$Z\big|_{a=0}\begin{Bmatrix} > \\ = \\ < \end{Bmatrix}0 \Leftrightarrow a\begin{Bmatrix} < \\ = \\ > \end{Bmatrix}\hat{a} \qquad \text{(A5).}$$

Recall from Lemma 1, part (ii) that $D_E^{**} > D_I^{**} \Rightarrow m_2 = 0$. Hence, from (A4) $Z > 0 \Rightarrow m_2 = 0$. Considering $Z = 0$, we can prove by contradiction that this implies $m_2 = 0$. Suppose $Z = 0$ and $m_2 > 0$ hold at the same time. In this case, it follows from (A4) that $D_E^{**} > D_I^{**}$, contradicting $m_2 > 0$. Hence, $Z \geq 0 \Rightarrow m_2 = 0$ must be true.

From (A4) and (A5) it follows $a \leq \hat{a} \Leftrightarrow D_E^{**} \geq D_I^{**}$, with a strict inequality on the RHS implying and being implied by one on the LHS. Finally, consider $a > \hat{a}$. From (A5), this implies $Z\big|_{a=0} < 0$. For $a = 0$ and, therefore, for $t_E = t_I = 0$, (ICI) and (ICE) imply $\hat{U}(b_I{}^{**}, r_I{}^{**}, b_E) - \hat{U}(b_E{}^{**}, r_E{}^{**}, b_E) = \hat{U}(b_I{}^{**}, r_I{}^{**}, b_I) - \hat{U}(b_E{}^{**}, r_E{}^{**}, b_I) = 0$. However, it is readily checked from figure 2b that for the utility specification in (5), this is only satisfied for a pooling allocation, i.e. for $\{b_E{}^{**} = b_I{}^{**} = b_P; r_E{}^{**} = r_I{}^{**} = r_P\}$. Hence, $a > \hat{a} \Rightarrow D_E^{**} = D_I^{**}$.■

Corollary: As $Z\big|_{a=0} < 0$ and $D_E^{**} = D_I^{**}$ both hold for $a > \hat{a}$, it follows from (A4) that $a > \hat{a} \Rightarrow m_2\big|_{a=0} > 0$.■

Proof of Lemma 4: From the first-order conditions (9e) and (9f), respectively, it follows for $m_2 = 0$ that $\hat{H}_b(b_E, r_E, b_E) = 1 + \frac{m_1}{l}\hat{U}_b(b_E, r_E, b_I) = V_b^E > 0$ and $\hat{H}_r(b_E, r_E, b_E) = 1 + \frac{m_1}{l}\hat{U}_r(b_E, r_E, b_I) = V_r^E > 0$. As $\hat{U}_b(\cdot, b_I) > 0$, $\hat{U}_r(\cdot, b_I) > 0$ and $m_1 \geq 0$, we have $V_b^E \geq 1$ and $V_r^E \geq 1$. It follows from the second-order condition $\hat{H}_{bb}^E \hat{H}_{rr}^E - \left(\hat{H}_{br}^E\right)^2 > 0$ that the optimal values $\hat{b}_E\left(V_b^E, V_r^E\right)$ and $\hat{r}_E\left(V_b^E, V_r^E\right)$ are decreasing functions in V_b^E and V_r^E. But then $b_E^{**} = \hat{b}_E\left(V_b^E, V_r^E\right) \leq \hat{b}_E(1,1) = b_E^*$ and $r_E^{**} = \hat{r}_E\left(V_b^E, V_r^E\right) \leq \hat{r}_E(1,1) = r_E^*$ is always true. By a similar proof, it can be shown that $b_I^{**} \geq b_I^*$ and $r_I^{**} \geq r_I^*$. Together this implies $B_I^{**} \geq B_I^*$ and $B_E^{**} \leq B_E^*$. This proves the relevant inequalities in parts (i)-(iii).

Note that the budget lines in (r,b) space (see figure 2a) have the slope –1. As indifference curves are strictly convex, it follows for $D_E \geq D_I$ that $\left.\frac{db(b_I, r_I)}{dr}\right|_{\hat{U}(b_E, r_E, b_I) = \hat{U}(b_I, r_I, b_I)} = \frac{-(r-a)b_I}{(a+b_I)r_I} \leq -1$ is sufficient for $B_I < B_E$. Using the definition $\hat{D}_I = \frac{a+b_I}{r-a}$, the condition can be equally expressed as $D_I \geq \hat{D}_I$. Similarly, $\left.\frac{db(b_E, r_E)}{dr}\right|_{\hat{U}(b_E, r_E, b_I) = \hat{U}(b_I, r_I, b_I)} = \frac{-(r-a)b_E}{(a+b_I)r_E} \geq -1$ or equivalently $D_E \leq \hat{D}_I$ are sufficient for $B_I > B_E$.

Consider now $\boldsymbol{a} \leq 0$, where $m_2 = 0$. It then follows from (9c)-(9f) that (10b) can be rewritten to $D_I^{**} = \hat{H}_b(\cdot, \boldsymbol{b}_E)\left(D_I^* - \hat{D}_I\right) + \hat{D}_I$. Hence, $\boldsymbol{a} \leq 0 \Leftrightarrow D_I^* \geq \hat{D}_I \Leftrightarrow D_I^{**} \geq \hat{D}_I \Rightarrow B_I^{**} < B_E^{**}$. This proves part (i). Similarly, for $\boldsymbol{a} \in [\overline{\boldsymbol{a}}, \hat{\boldsymbol{a}}[$ and thus $m_2 = 0$, we obtain $D_E^{**} = \hat{H}_b(\cdot, \boldsymbol{b}_E)\left(D_E^* - \hat{D}_I\right) + \hat{D}_I$. Here, $\boldsymbol{a} \geq \overline{\boldsymbol{a}} \Leftrightarrow D_E^* \leq \hat{D}_I \Leftrightarrow D_E^{**} \leq \hat{D}_I \Rightarrow B_E^{**} < B_I^{**}$. This proves part (ii). Finally, for $\boldsymbol{a} \geq \hat{\boldsymbol{a}}$ a pooling allocation obtains where $D_E^{**} = D_I^{**} = D_P \Leftrightarrow B_E^{**} = B_I^{**} = B_P$, which proves part (iii).■

Proof of Lemma 5: In what follows we write the equilibrium degree of delegation $D_i^{**} = d^i(\boldsymbol{a}, a),\ i = E, I$ and the shadow prices $m_1 = \hat{m}(\boldsymbol{a}, a)$ and $m_2 = \hat{\hat{m}}(\boldsymbol{a}, a)$ as functions of the parameters $\boldsymbol{a}$ and a. From (9a), we obtain $m_3 = \max\left\{(1-\boldsymbol{l}) - a\left|\hat{m}(\boldsymbol{a}, a) - \hat{\hat{m}}(\boldsymbol{a}, a)\right|, 0\right\}$.

(i) Consider $\boldsymbol{a} < \hat{\boldsymbol{a}}$. This implies $d^I(\boldsymbol{a}, a) < d^E(\boldsymbol{a}, a)$. Here, $d^I(\boldsymbol{a}, 0) < d^E(\boldsymbol{a}, 0)$ was shown in Lemma 3, part (iii). We can now prove $d^I(\boldsymbol{a}, a) < d^E(\boldsymbol{a}, a)$ by contradiction. Suppose $d^I(\boldsymbol{a}, a) > d^E(\boldsymbol{a}, a)$. As we see below, two cases can arise for $a > 0$. First, the principal sets $t_I = 0$. Here, the incentive problem is identical to the case $a = 0$, implying that $d^i(\boldsymbol{a}, a) = d^i(\boldsymbol{a}, 0)$ and, thus, a contradiction. Second, $t_I > 0$. As the transfer is costly, it is used in addition to the instruments $\{B_i, D_i\}$ only if this relaxes the incentive constraint (ICI). This implies $\left|d^i(\boldsymbol{a}, a) - D_i^*\right| \leq \left|d^i(\boldsymbol{a}, 0) - D_i^*\right|$. Since $D_I^* < d^I(\boldsymbol{a}, 0) < d^E(\boldsymbol{a}, 0) < D_E^*$, it must then be true that $d^I(\boldsymbol{a}, a) < d^I(\boldsymbol{a}, 0) < d^E(\boldsymbol{a}, 0) < d^E(\boldsymbol{a}, a)$, again a contradiction.

We can now show that there exists a function $\underline{a}(\boldsymbol{a}) \in (0, \infty)$ such that $m_3 > 0 \Leftrightarrow a \leq \underline{a}(\boldsymbol{a})$. Since $d^I(\boldsymbol{a}, a) < d^E(\boldsymbol{a}, a)$, this implies $m_2 = 0$ and $m_3 = \max\{(1-\boldsymbol{l}) - a\hat{m}(\boldsymbol{a}, a), 0\}$. For $m_3 > 0$ comparative static analysis of the system $\hat{U}(b_I, r_I, \boldsymbol{b}_I) - \hat{U}(b_E, r_E, \boldsymbol{b}_I) = 0$, (9a) and (9c)-(9f) with $m_2 = 0$ yields $\hat{m}_a(\boldsymbol{a}, a) = 0$. Hence, $\left.\frac{d[(1-\boldsymbol{l}) - a\hat{m}(\boldsymbol{a}, a)]}{da}\right|_{m_3 > 0} = -\hat{m}(\boldsymbol{a}, a) < 0$. Since $m_3 > 0 \Rightarrow \hat{m}_a = 0$, it follows that $m_3 > 0 \Rightarrow \hat{m}(\boldsymbol{a}, a) = \hat{m}(\boldsymbol{a}, 0) = \hat{m}(\boldsymbol{a})$. Furthermore, $\hat{m}(\boldsymbol{a}) \in (0, \infty)$ is positive and finite. But then $\lim_{a \to 0}[(1-\boldsymbol{l}) - a\hat{m}(\boldsymbol{a})] = (1-\boldsymbol{l}) > 0$ and

$\lim_{a\to\infty}[(1-l)-a\hat{m}(a)]=-\infty$. Hence, there exists a function $\underline{a}(a)\in(0,\infty)$ such that $m_3=(1-l)-a\hat{m}(a)>0 \Leftrightarrow a\le\underline{a}(a)$.

(ii) Consider $a\ge\hat{a}$. For this case we can show that on the domain $[\hat{a},\infty)$ there exists (a) a unique function $\overline{a}(a)\in(0,\infty)$ such that $m_3>0\Leftrightarrow a\le\overline{a}(a)$; (b) a unique correspondence $\hat{a}(a)\in(0,\infty)$ such that $m_2=\hat{\hat{m}}(a,a)>0\Leftrightarrow a\le\hat{a}(a)$; and (c) a unique function $\hat{\hat{a}}(a)\in(0,\infty)$ such that $d^I(a,a)\ge d^E(a,a)\Leftrightarrow a\le\hat{\hat{a}}(a)$; where

(iii) $\overline{a}(a)\le\hat{a}(a)\le\hat{\hat{a}}(a)$ and $\underline{a}(\hat{a})=\overline{a}(\hat{a})=\hat{a}(\hat{a})=\hat{\hat{a}}(\hat{a})$.

For convenience, we prove part (iii) first, assuming the existence of $\overline{a}(a)$, $\hat{a}(a)$ and $\hat{\hat{a}}(a)$. Consider $a>\hat{a}$. By way of contradiction suppose $\hat{\hat{a}}(a)<\hat{a}(a)$ and consider $a\in\left]\hat{\hat{a}}(a),\hat{a}(a)\right[$. Here, $a<\hat{a}(a)\Rightarrow m_2>0$ and $a>\hat{\hat{a}}(a)\Rightarrow d^I(a,a)<d^E(a,a)$. We have shown in part (iii) of Lemma 1 that this is a contradiction. Hence, $\hat{\hat{a}}(a)>\hat{a}(a)$ must be true. Next, by way of contradiction suppose $\hat{a}(a)<\overline{a}(a)$ and consider $a\in\left]\hat{a}(a),\overline{a}(a)\right[$. Here, $a>\hat{a}(a)\Rightarrow m_2=0$, whereas $a<\overline{a}(a)\Rightarrow m_3>0\Rightarrow t_I=0$. This implies $d^i(a,a)=d^i(a,0)$. But then, from Lemma 3, part (iv) we know that $a>\hat{a}$ implies $d^E(a,a)=d^E(a,0)=d^I(a,0)=d^I(a,a)$. We have shown as corollary to the proof of Lemma 3 that $d^I(a,0)=d^E(a,0)$ and $a>\hat{a}$ imply $m_2>0$. Hence, a contradiction and $\hat{a}(a)>\overline{a}(a)$ must be true.

Now consider $a=\hat{a}$. Suppose $a<\overline{a}(a)\Rightarrow m_3>0\Rightarrow t_I=0$ and therefore $d^E(a,a)=d^E(a,0)=d^I(a,0)=d^I(a,a)$. We have shown as corollary to Lemma 3 that $d^I(a,0)=d^E(a,0)$ and $a=\hat{a}$ imply $m_2=0$. By definition, $a>\overline{a}(a)\Rightarrow m_3=0\Rightarrow t_I>0$. As we have argued before, given $m_2=0$, the use of financial transfers implies a relaxation of (ICI) and therefore, $d^I(\hat{a},a)<d^I(\hat{a},0)=d^E(\hat{a},0)<d^E(\hat{a},a)$ which is still consistent with $m_2=0$. But then, it follows that $\overline{a}(\hat{a})=\hat{\hat{a}}(\hat{a})$. Since $\overline{a}(\hat{a})\le\hat{a}(\hat{a})\le\hat{\hat{a}}(\hat{a})$ must also be true, it follows that $\overline{a}(\hat{a})=\hat{a}(\hat{a})=\hat{\hat{a}}(\hat{a})$. Finally, $\lim_{a\to\hat{a}^+}\left\{(1-l)-a\left[\hat{m}(a,a)-\hat{\hat{m}}(a,a)\right]\right\}=(1-l)-a\hat{m}[\hat{a},a]=\lim_{a\to\hat{a}^-}\{(1-l)-a\hat{m}(a,a)\}$ implies $\underline{a}(\hat{a})=\overline{a}(\hat{a})$.

To prove part (ii) (c) reconsider $a>\hat{a}$. For $a\to\infty$, separation can be implemented at the first-best values $\{b_E^*,r_E^*,b_I^*,r_I^*\}$ for an infinitesimally small value of t_I. This implies $d^E(a,\infty)>d^I(a,\infty)$. Furthermore, comparative statics for the system $\hat{U}(b_I,r_I,b_I)-\hat{U}(b_E,r_E,b_I)+at_I=0$, (9a) and (9c)-(9f), where $m_2=0$, yield after tedious calculations $\left.(d_a^E-d_a^I)\right|_{d^E(a,a)=d^I(a,a)}>0$. Together with the fact that $d^E(a,0)=d^I(a,0)$ for $a>\hat{a}$, this implies the existence of a unique $\hat{\hat{a}}(a)\in(0,\infty)$ such that $d^I(a,a)\ge d^E(a,a)\Leftrightarrow a\le\hat{\hat{a}}(a)$.

To prove part (ii) (b), observe that $\hat{\hat{m}}(a,\infty)=0$. Furthermore, we have shown as part of the proof of Lemma 3, part (iv) that $a>\hat{a} \Leftrightarrow \hat{\hat{m}}(a,0)>0$. Finally, comparative static analysis for the system $\hat{U}(b_I,r_I,b_I)-\hat{U}(b_E,r_E,b_I)+at_I=0$, $\hat{U}(b_E,r_E,b_E)-\hat{U}(b_I,r_I,b_E)-at_I=0$, (9a) and (9c)-(9f) yields after tedious calculations $\hat{\hat{m}}_a(a,a)\big|_{m_2=0}<0$. It follows that there exists a unique $\hat{a}(a)\in(0,\infty)$ such that $m_2=\hat{\hat{m}}(a,a)>0 \Leftrightarrow a\le\hat{a}(a)$.

To prove part (ii) (a), consider $m_3=\max\left\{(1-l)-a\left[\hat{m}(a,a)-\hat{\hat{m}}(a,a)\right],0\right\}$. Comparative static analysis for the full pooling system $\hat{U}(b_I,r_I,b_I)-\hat{U}(b_E,r_E,b_I)=0$, $\hat{U}(b_E,r_E,b_E)-\hat{U}(b_I,r_I,b_E)=0$, (9a) and (9c)-(9f) with $m_2>0$ yields $\hat{m}_a(a,a)=\hat{\hat{m}}_a(a,a)=0$. Hence, $\frac{d\left[(1-l)-a\left[\hat{m}(a,a)-\hat{\hat{m}}(a,a)\right]\right]}{da}\bigg|_{m_3>m_2>0}=-\left[\hat{m}(a,a)-\hat{\hat{m}}(a,a)\right]<0$. Since $m_3>0\Rightarrow\hat{m}_a=\hat{\hat{m}}_a=0$, we have $m_3>0\Rightarrow\hat{m}(a,a)-\hat{\hat{m}}(a,a)=\hat{m}(a,0)-\hat{\hat{m}}(a,0)=\hat{m}(a)-\hat{\hat{m}}(a)$. Furthermore, $\hat{m}(a)-\hat{\hat{m}}(a)\in\,]0,\infty[$. But then $\lim_{a\to0}\left\{(1-l)-a\left[\hat{m}(a)-\hat{\hat{m}}(a)\right]\right\}=(1-l)>0$ and $\lim_{a\to\infty}\left\{(1-l)-a\left[\hat{m}(a)-\hat{\hat{m}}(a)\right]\right\}=-\infty$. Hence, there exists a value $\overline{a}(a)\in(0,\infty)$ such that $m_3=(1-l)-a\left[\hat{m}(a)-\hat{\hat{m}}(a)\right]>0\Rightarrow a\le\overline{a}(a)$.■

Proof of Corollary L5:

(i) $t_I^{**}=0$ in areas I and V follows from definition of $\underline{a}(a)$ and $\overline{a}(a)$ and $m_3>0\Leftrightarrow t_I=0$.

(ii) $D_E^{**}>D_I^{**}$ in areas I and II follows from $a<\hat{a}\Rightarrow D_E^{**}>D_I^{**}$ as shown for part (i) of Lemma 5 together with the definition of $\hat{\hat{a}}(a)$ implying $d^I(a,a)<d^E(a,a)\Leftrightarrow a>\hat{\hat{a}}(a)$. $D_E^{**}=D_I^{**}$ in area V follows from the fact that $t_I^{**}=0$ and $a\ge\hat{a}$ imply full pooling. $D_E^{**}<D_I^{**}$ in areas III and IV follows as a residual.■

Proof of Proposition 4: For the moment, we conjecture that both moral hazard constraints (MHE) and (MHI) are binding and verify this later on (see part i). Inserting $t_i=\frac{1}{a}\left[U\left(e_i^*,b_i,r_i,b_i\right)-U\left(\overline{e}_i,b_i,r_i,b_i\right)\right]$, $i=E,I$ into (2) and maximising with respect to $\{\overline{e}_i,b_i,r_i\}$, $i=E,I$ provides the first-order conditions

$$h(b_i,r_i,b_i)-\tfrac{1}{a}U_e\left(\overline{e}_i,b_i,r_i,b_i\right)=\tfrac{1}{a}\left\{\left[a+u(b_i,r_i)\right]h(b_i,r_i,b_i)-\overline{e}_i\right\}=0 \qquad \text{(A6a);}$$

$$\overline{e}_i h_x(b_i,r_i,b_i)-\tfrac{1}{a}\left[U_x\left(e_i^*,b_i,r_i,b_i\right)+U_e\left(e_i^*,b_i,r_i,b_i\right)\hat{e}_x(b_i,r_i,b_i)-U_x\left(\overline{e}_i,b_i,r_i,b_i\right)\right]-1$$

$$=\overline{e}_i h_x(b_i,r_i,b_i)+\tfrac{1}{a}\left(\overline{e}_i-e_i^*\right)\begin{bmatrix}u(b_i,r_i)h_x(b_i,r_i,b_i)\\+u_x(b_i,r_i)h(b_i,r_i,b_i)\end{bmatrix}-1=0;\quad i=E,I;\quad x=b,r \qquad \text{(A6b);}$$

where $U_e\left(e_i^*,b_i,r_i,b_i\right)=0$, $U_e\left(\overline{e}_i,b_i,r_i,b_i\right)=u(\cdot)h(\cdot)-\overline{e}_i$ and $U_x\left(e_i,b_i,r_i,b_i\right)=e_i\left[u(\cdot)h_x(\cdot)+u_x(\cdot)h(\cdot)\right]$; $x=b,r$. The following can now be derived.

(i) Using (A6a) and (3) we can rewrite $\bar{e}_i = [a + u(b_i, r_i)]h(b_i, r_i, b_i) = ah(b_i, r_i, b_i) + e_i^* > e_i^*$. Inserting $\bar{e}_i$ and e_i^* into the moral hazard constraints (MHi) and cancelling terms yields $t_i = \frac{a}{2} h(b_i, r_i, b_i)^2 > 0$.

(ii) Tedious comparative static analysis on the system (A6a) and (A6b) provides $\frac{db_i^+}{db_i} \geq 0$ and $\frac{dr_i^+}{db_i} \geq 0$ and $\frac{d\bar{e}_i^+}{db_i} \geq 0$. Hence, $b_E^+ \geq b_I^+$, $r_E^+ \geq r_I^+$, implying $t_E^+ = \frac{a}{2} h(b_E^+, r_E^+, b_E)^2 > \frac{a}{2} h(b_I^+, r_I^+, b_I)^2 = t_I^+$, and $\bar{e}_E^+ \geq \bar{e}_I^+$.

(iii) Using (3) we can rewrite $u(b_i, r_i)h_x(b_i, r_i, b_i) + u_x(b_i, r_i)h(b_i, r_i, b_i) = \hat{e}_x(b_i, r_i, b_i);\ x = b, r$. Substituting this and $\bar{e}_i = ah(b_i, r_i, b_i) + e_i^*$ into (A6b) allows us to rewrite the first-order conditions to $\bar{e}_i h_x(\cdot) + h(\cdot)\hat{e}_x - 1 = 0;\quad i = E, I;\quad x = b, r$. Inserting into this $\bar{e}_i = [a + u(b_i, r_i)]h(b_i, r_i, b_i)$, $h_b(\cdot) = b_i h(\cdot) b_i^{-1}$, $h_r(\cdot) = rh(\cdot) r_i^{-1}$, $\hat{e}_b(\cdot) = (b_i + a)u(\cdot)h(\cdot)b_i^{-1}$ and $\hat{e}_r(\cdot) = (r - a)u(\cdot)h(\cdot)r_i^{-1}$ and solving for b_i and r_i, respectively, gives $b_i^+ = [b_i a + (2b_i + a)u(\cdot)]h(\cdot)^2$ and $r_i^+ = [ra + (2r - a)u(\cdot)]h(\cdot)^2$. Hence, $D_i^+ = \frac{b_i^+}{r_i^+} = \frac{b_i a + (2b_i + a)u(\cdot)}{ra + (2r - a)u(\cdot)}$. It is then readily verified that $D_i^+ > D_i^* = \frac{2b_i + a}{2r - a} \Leftrightarrow a < 0$.■

Proof of Proposition 5:

(i) Follows directly from setting $m_1^E = m_1^I = 0$ in (16c) and (16d).

Using (16a) and (16b), where $m_1^E = m_1^I = 0$, we can rewrite (16e/f) and (16g/h) as

$$\left\langle \begin{matrix} (1-l)\{\bar{e}_{II} h_x(b_I, r_I, b_I) + \frac{1}{a}[U_x(\bar{e}_{II}, b_I, r_I, b_I) - U_x(e_{II}^*, b_I, r_I, b_I)] - 1\} \\ + m_2^I U_x(e_{II}^*, b_I, r_I, b_I) - m_2^E U_x(e_{EI}^*, b_I, r_I, b_E) \end{matrix} \right\rangle = 0;\ x = b, r \qquad \text{(16e'/f')},$$

$$\left\{ \begin{matrix} l\{\bar{e}_{EE} h_x(b_E, r_E, b_E) + \frac{1}{a}[U_x(\bar{e}_{EE}, b_E, r_E, b_E) - U_x(e_{EE}^*, b_E, r_E, b_E)] - 1\} \\ + m_2^E U_x(e_{EE}^*, b_E, r_E, b_E) - m_2^I U_x(e_{IE}^*, b_E, r_E, b_I) \end{matrix} \right\} = 0;\ x = b, r \qquad \text{(16g'/h')}.$$

It can then be shown that the optimum entails (a) $m_2^I > m_2^E \geq 0$; (b) $m_2^E = 0$ if $D_E > D_I$ or if $D_E = D_I$ and $B_E \neq B_I$.

(a) Working with (16e'/f') and (16g'/h'), the proof is analogous to the proof of part (i) of Lemma 1.

(b) We prove this by contradiction. Thus, suppose $m_2^E > 0$ when $D_E > D_I$. Recall that $m_2^E > 0 \Rightarrow U(e_{EI}^*, B_I, D_I, b_E) \geq U(e_{EE}^*, B_E, D_E, b_E)$. Using figure 2a), where E corresponds to the allocation $\{B_E, D_E\}$, we note that this implies an allocation $\{B_I', D_I\}$ on the D_I-ray to the north-east of its intersection with the $\hat{U}^E$ indifference curve. But then

$U\left(e_{II}^{*}, B_I', D_I, b_I\right) > U\left(e_{IE}^{*}, B_E, D_E, b_E\right) = \hat{U}^I$. This contradicts $m_2^I > m_2^E \geq 0$. By a similar argument, it can be shown that $m_2^E > 0$ contradicts $D_E = D_I$ unless $B_E = B_I$.

From (16e'/f') and (16g'/h'), respectively, one obtains after tedious calculations

$$D_I^{++} = D_I^{+} + \frac{m_2^I \hat{U}_r(b_I, r_I, b_I)}{1-l}\left(\hat{D}_I - D_I^{+}\right) + \frac{m_2^E \hat{U}_r(b_I, r_I, b_E)}{1-l}\left(D_I^{+} - \hat{D}_E\right) \qquad \text{(A7a)},$$

$$D_E^{++} = D_E^{+} + \frac{m_2^I \hat{U}_r(b_E, r_E, b_I)}{l}\left(D_E^{+} - \hat{D}_I\right) + \frac{m_2^E \hat{U}_r(b_E, r_E, b_E)}{l}\left(\hat{D}_E - D_E^{+}\right) \qquad \text{(A7b)}.$$

with D_i^{+} defined by (14). Note that (A7a) and (A7b) are similar to (10a) and (10b).

(ii) From the first-order conditions in (16a) and (16b), where $m_1^E = m_1^I = 0$, we obtain $m_3^I = \max\left\{\frac{1-l}{a} - m_2^I, 0\right\}$ and $m_3^E = \max\left\{\frac{l}{a} - m_2^E, 0\right\}$. Tedious calculations show that $\left.\frac{dm_2^i}{da}\right|_{m_3^i > 0} = 0$, implying $\left.\frac{d}{da}\left(\frac{1-l}{a} - m_2^I\right)\right|_{m_3^I > 0} < 0$ and $\left.\frac{d}{da}\left(\frac{l}{a} - m_2^E\right)\right|_{m_3^E > 0} < 0$. Observing that m_2^i, $i = E, I$ take on finite non-negative values for $a = 0$, it follows in analogy to the proof of Lemma 5 part I that there exists $\overline{a}_i^{+}(a) \in (0, \infty)$; $i = E, I$ such that $m_3^i > 0$ if and only if $a \in \left[0, \overline{a}_i^{+}(a)\right]$. The transfers are then given by $t_i^{++} = \frac{1}{a}\left[U\left(e^*_{ii}, B_i, D_i, b_i\right) - U\left(\overline{e}_{ii}, B_i, D_i, b_i\right)\right]$, $i = E, I$ as indicated in the Proposition.

(iii) If $a < \overline{a}_i^{+}(a)$, $i = E, I$, then $U\left(\overline{e}_{ii}, b_i, r_i, b_i\right) + at_i = U\left(e_{ii}^{*}, b_i, r_i, b_i\right)$. The adverse selection constraints are then given by $U\left(e_{II}^{*}, b_I, r_I, b_I\right) \geq U\left(e_{IE}^{*}, b_E, r_E, b_I\right)$ and $U\left(e_{EE}^{*}, b_E, r_E, b_E\right) \geq U\left(e_{EI}^{*}, b_I, r_I, b_E\right)$. This is equivalent to the case $t_i = 0$, $i = E, I$ in the setting without contractible output. In analogy to Lemma 5 together with figure 3 it is then easy to establish a pooling equilibrium (area V). Using an argument similar to the one used in the proof of parts (iii) and (iv) of Lemma 3, we find from (A7a) and (A7b)

$$D_E^{++} = D_I^{++} \Leftrightarrow \hat{Z} + m_2^E\left[\frac{\hat{U}_r(b_E, r_E, b_E)}{l} + \frac{\hat{U}_r(b_I, r_I, b_E)}{1-l}\right]\left(\hat{D}_E - \hat{D}_I\right) = 0,$$

where

$$\hat{Z} := \begin{Bmatrix} \left[\overline{e}_{EE} h_r(b_E, r_E, b_E) + h(b_E, r_E, b_E)\hat{e}_r(b_E, r_E, b_E)\right]\left(D_E^{+} - \hat{D}_I\right) \\ -\left[\overline{e}_{II} h_r(b_I, r_I, b_I) + h(b_I, r_I, b_I)\hat{e}_r(b_I, r_I, b_I)\right]\left(D_I^{+} - \hat{D}_I\right) \end{Bmatrix} \qquad \text{(A8)}.$$

Hence, for $m_2^E \geq 0$, $D_E^{++} = D_I^{++} \Leftrightarrow \hat{Z} \leq 0$. Assuming pooling, i.e. $b_E^{++} = b_I^{++} = b_P$ and $r_E^{++} = r_I^{++} = r_P$ one finds from (A8) that $\left.\hat{Z}\right|_{\{b_E^{++} = b_I^{++} = b_P; r_E^{++} = r_I^{++} = r_P\}} \leq 0 \Leftrightarrow b_P^{2b_E}\left(D_E^{+} - \hat{D}_I\right) - b_P^{2b_I}\left(D_I^{+} - \hat{D}_I\right) \leq 0$. Inserting from (6) and (14) while observing $D_E^{++} = D_I^{++}$ one finds

$$\hat{Z}\Big|_{\{b_E{}^{++}=b_I{}^{++}=b_P;r_E{}^{++}=r_I{}^{++}=r_P\}} \leq 0 \Leftrightarrow \begin{Bmatrix} -aa\left[(r+b_E)b_P^{2b_E}-(r+b_I)b_P^{2b_I}\right] \\ +u(D_P)(\hat{a}-a)\left[(r+2b_E-b_I)b_P^{2b_E}-(r+b_I)b_P^{2b_I}\right] \end{Bmatrix} \leq 0,$$

where $\hat{a}$ as defined in (12). As the first term in bracelets is unambiguously negative, it follows that $a \geq \hat{a} \Rightarrow \hat{Z}\Big|_{\{b_E{}^{++}=b_I{}^{++}=b_P;r_E{}^{++}=r_I{}^{++}=r_P\}} < 0$. But then $a \geq \hat{a}$ and $a < \overline{a}_i^{\,+}(a)$ are sufficient for pooling. Further boundaries corresponding to those in figure 3 can be derived.■

Proof of Lemma 8: From (15a) and (15b), we find that case 1 applies if and only if $U(e_{IE}^*, b_E, r_E, b_I) > U(\overline{e}_{IE}, b_E, r_E, b_I) + at_E$ and $U(e_{EE}^*, b_E, r_E, b_E) > U(\overline{e}_{EI}, b_I, r_I, b_E) + at_I$. Inserting $t_E = \frac{1}{a}\left[U(e_{EE}^*, b_E, r_E, b_E) - U(\overline{e}_{EE}, b_E, r_E, b_E)\right]$ and $t_I = \frac{1}{a}\left[U(e_{IE}^*, b_E, r_E, b_I) - U(\overline{e}_{II}, b_I, r_I, b_I)\right]$, these conditions can be written as

$$U(\overline{e}_{EE}, b_E, r_E, b_E) - U(\overline{e}_{IE}, b_E, r_E, b_I) > U(e_{EE}^*, b_E, r_E, b_E) - U(e_{IE}^*, b_E, r_E, b_I) \Leftrightarrow \overline{e}_{EE} > e_{IE}^* \quad \text{(113a)};$$

$$U(e_{EE}^*, b_E, r_E, b_E) - U(e_{IE}^*, b_E, r_E, b_I) > U(\overline{e}_{EI}, b_I, r_I, b_E) + U(\overline{e}_{II}, b_I, r_I, b_I)$$

$$\Leftrightarrow \overline{e}_{II}^2 < \frac{u(b_E, r_E)^2\, h(b_I, r_I, b_E)^2 \left[h(b_E, r_E, b_E)^2 - h(b_E, r_E, b_I)^2\right]}{\left[h(b_I, r_I, b_E)^2 - h(b_I, r_I, b_I)^2\right]} = \left(e_{IE}^*\right)^2 v \ ,$$

where $v := \frac{U(e_{EE}^*, b_E, r_E, b_E) - U(e_{IE}^*, b_E, r_E, b_I)}{U(e_{EI}^*, b_I, r_I, b_E) - U(e_{II}^*, b_I, r_I, b_I)} \geq 1$ follows from the monotonicity condition. It is readily verified that the RHS inequality in (113a) is always true, leaving the condition as stated in the Lemma.■

Proof of Lemma 9: From (15a) $m_1^I = 0$ if and only if

$$U(\overline{e}_{EE}, b_E, r_E, b_E) - U(\overline{e}_{IE}, b_E, r_E, b_I) > U(\overline{e}_{EI}, b_I, r_I, b_E) - U(\overline{e}_{II}, b_I, r_I, b_I)$$

$$\Leftrightarrow \left\{\left[\tfrac{h(b_E, r_E, b_E)}{h(b_E, r_E, b_I)}\right]^2 - 1\right\} \overline{e}_{EE}{}^2 > \left\{1 - \left[\tfrac{h(b_I, r_I, b_I)}{h(b_I, r_I, b_E)}\right]^2\right\} \overline{e}_{II}{}^2 \ .■$$

Proof of Proposition 6:

(i) Follows immediately from (16c) and (16d), where $m_1^E > m_1^I = 0$. Noting that $U_e(\overline{e}_{EI}, b_I, r_I, b_E) - U_e(\overline{e}_{II}, b_I, r_I, b_I) = \overline{e}_{II}\left\{1 - \left[\tfrac{h(b_I, r_I, b_I)}{h(b_I, r_I, b_E)}\right]^2\right\} > 0$ it follows from (16c) that $\overline{e}_I^{+++}$ is distorted downwards.

(ii) Follows immediately from $m_1^E > 0$ and $m_2^I > 0$. The latter is true since (16e) implies $b_I^{+++} < b_I^+$. But $m_2^I > 0$ follows in analogy to the proof of part (i) of Lemma 1.

(iii) Follows immediately from (16g/h), where $m_1^I = 0$.

(iv) Follows from (16e/f), where $m_1^E > 0$. Here, $U_b(\overline{e}_{EI}, b_I, r_I, b_E) - U_b(\overline{e}_{II}, b_I, r_I, b_I) = -\overline{e}_{II}^2 \frac{h(b_I, r_I, b_I)}{h(b_I, r_I, b_E)^3} \Lambda > 0$, where

$\Lambda := [h_b(b_I, r_I, \boldsymbol{b}_I)h(b_I, r_I, \boldsymbol{b}_E) - h_b(b_I, r_I, \boldsymbol{b}_E)h(b_I, r_I, \boldsymbol{b}_I)] = (\boldsymbol{b}_I - \boldsymbol{b}_E)b_I^{b_I + b_E - 1} r_I^{2r} < 0$ and $U_r(\bar{e}_{EI}, b_I, r_I, \boldsymbol{b}_E) - U_r(\bar{e}_{II}, b_I, r_I, \boldsymbol{b}_I) = 0$. Hence, it follows from (16e/f) that b_I^{+++} is distorted downwards, that r_I^{+++} remains undistorted, and, by implication, that B_I^{+++} and D_I^{+++} are distorted downwards.■

Proof of Proposition 7: In analogy to the analysis set out in section 4 the degree of delegation under asymmetric information can be determined as

$$D_L^{**} = D_L^{*} + \tfrac{m_1 \hat{U}_r(b_L, r_L, a_L)}{1-l}\left(\hat{D}_L - D_L^{*}\right) + \tfrac{m_2 \hat{U}_r(b_L, r_L, a_H)}{l}\left(D_L^{*} - \hat{D}_H\right);$$

$$D_H^{**} = D_H^{*} + \tfrac{m_1 \hat{U}_r(b_H, r_H, a_L)}{1-l}\left(D_H^{*} - \hat{D}_L\right) + \tfrac{m_2 \hat{U}_r(b_H, r_H, a_H)}{l}\left(\hat{D}_H - D_H^{*}\right).$$

with $\hat{D}_i = \frac{b + a_i}{r - a_i}$ and $D_i^{*} = \frac{2b + a_i}{2r - a_i}$, $i = H, L$; where $m_1 > m_2 \geq 0$ and $D_H^{**} > D_L^{**} \Rightarrow m_2 = 0$. Using the above expressions, the following is readily established.

(i) $D_L^{**} > D_L^{*} \Leftrightarrow \hat{D}_L > D_L^{*} \Leftrightarrow \boldsymbol{a}_L > 0$;

(ii) $D_H^{**} < D_H^{*} \Leftrightarrow D_H^{*} < \hat{D}_L \Leftrightarrow \boldsymbol{a}_H < 2\boldsymbol{a}_L$

(iii) $D_H^{**} = D_L^{**} = D_P \Leftrightarrow X(\boldsymbol{a}_L) = (\boldsymbol{a}_H - \boldsymbol{a}_L)D_P^{a_H} + \boldsymbol{a}_L\left(D_P^{a_L} - D_P^{a_H}\right) \leq 0$.

Noting that $X(\tfrac{1}{2}\boldsymbol{a}_H) > 0$ and $X(\boldsymbol{a}_H) = 0$, the above inequality is true for a range $\boldsymbol{a}_L \in [\underline{\boldsymbol{a}}, \boldsymbol{a}_H]$, with $\underline{\boldsymbol{a}} \in (\tfrac{1}{2}\boldsymbol{a}_H, \boldsymbol{a}_H]$, if $\lim_{a_L \to a_H} X'(\boldsymbol{a}_L) = -1 + \boldsymbol{a}_H\left(\boldsymbol{a}_H D_H^{*\,-1} \tfrac{dD_P}{da_L} + \ln D_H^{*}\right) > 0$. Observing $\tfrac{dD_P}{da_L} \geq 0$, it follows that $\lim_{a_L \to a_H} X'(\boldsymbol{a}_L) = -1 + \boldsymbol{a}_H \ln D_H^{*} > 0$ is sufficient, thus, implying the condition provided in the Lemma.■