University of Heidelberg

Department of Economics

Discussion Paper Series | No. 505

Hidden Action, Identification, and Organization Design

Wendelin Schnedler

November 2010

Hidden Action, Identification, and Organization Design

Wendelin Schnedler*
Department of Economics
University of Heidelberg

First Version: September 2008
This Version: November 2010

Abstract

Incentives often fail in inducing economic agents to engage in a desirable activity; implementability is restricted. What restricts implementability? When does re-organization help to overcome this restriction? This paper shows that any restriction of implementability is caused by an *identification problem*. It also describes organizations that can solve this identification problem and provides conditions under which such organisations exist. Applying the findings to established and new moral hazard models yields insights into optimal organization design, uncovers the reason why certain organization designs, such as advocacy or specialization, overcome restricted implementability, and formalizes a wide-spread type of multi-tasking problem.

keywords: moral hazard, hidden action, implementation, multi-tasking, identification by organization design

JEL-Codes: D86, M52, J33, D82, M41

*wendelin.schnedler@awi.uni-heidelberg.de, Heidelberg University, Department of Economics, Bergheimer Straße 58, D-69115 Heidelberg, Germany

Arguably, one of the most important roles of a manager (be it in the private or public sector) is to ensure that production occurs in a desired manner and that policies are implemented as intended. To this end, the manager relies on agents (employees, public servants, production units, departments) who take decisions, e.g., on how much to work or on what input to use. Typically, agents' activities cannot be observed directly but have to be deduced from signals, which may be manipulated, influenced by others, or affected by factors beyond agents' control. These signals can be used to provide incentives and influence the activity. In numerous examples, however, the imperfect nature of signals means that incentives cannot induce the desired activity.[1]

Anecdotal evidence as well as theoretical models suggest that in these cases organization design, in particular the partition of tasks amongst agents, can help to obtain those activities.[2] A prominent example is Dewatripont and Tirole's advocates model (1999). In this model, two tasks conflict: exerting effort to find incriminating evidence increases the probability of conviction while searching for exonerating evidence reduces it. A single investigator cannot be motivated to do both, but with advocates, this becomes possible. The effect of organization is usually attributed to the conflicting nature of tasks (see e.g. Dewatripont, Jewitt, and Tirole, 2000). However, Ratto and Schnedler (2008) provide an example in which tasks do not conflict: the stability of a computer program increases with programming and debugging effort. Only the program's stability is observable but the manager prefers if stability is achieved by better programming rather than by a lot of debugging, e.g., because this facilitates maintenance at a later stage. With a single agent, who prefers debugging, the desired emphasis on programming cannot be induced. With two specialists, a programmer and a debugger, any activity can be implemented. Task conflict is thus not necessary for organization design to overcome restricted implementability and induce desired activities. So what restricts implementability? When does organization design help to overcome this restriction, and if so, how?

The main finding is the following. Restricted implementability is driven by a specific signal imperfection, which is best described by adapting an idea

[1]A classical example is AT&T, who paid programmers by the number of code lines to increase productivity but only obtained longer code. Similar examples can be found in the survey articles on incentives by Gibbons (1998) and Prendergast (1999).

[2]For example, Ciba Vision achieved its goal to generate more innovative products by separating research for new contact lens products from research to improve existing products (see "The Ambidextrous Organization" by O'Reilly and Thushman in Harvard Business Review April 2004).

from econometrics: any activity choice can be implemented if and only if agents' choices are *(conditionally) identified*. The relevant notion of identification is considerably weaker than that in econometrics. It is not necessary to deduce the whole activity from signals, rather it suffices that each agent's behavior could theoretically be inferred if the behavior of other agents were known.

Organization designs such as Dewatripont and Tirole's advocacy can overcome restricted implementability precisely because this weaker form of identification suffices. To see why, re-consider the program stability example. A single employee can produce a given stability in many ways. He will select the for him cheapest way, which is not necessarily the one desired by the manager. Accordingly, implementability is restricted. Now take two specialists, a programmer and a debugger, and suppose that the debugger engages in the desired amount of debugging. Then, the programmer can only produce a given stability in one specific way. In particular, the stability associated with the desired choices can only be generated if the programmer exerts the desired programming effort. The programmer's activity can be deduced from the stability *given* the debugger's choice and vice versa. Both agents' activity is (conditionally) identified. As a result, it becomes possible to attach incentives to stability outcomes such that the desired effort is the programmer's best response if the debugger carries out the desired amount of debugging. For example, the programmer's marginal rewards can be adjusted such that they equal his marginal costs if and only if he exerts the desired effort. Similarly, one can ensure that the desired amount of debugging is a best response to the desired programming effort. As a result, desired activity choices form a Nash equilibrium.

The main result is based on two propositions that are interesting in their own right. First, an activity can be implemented if and only if the activity's marginal effect on agent's costs are a linear combination of that on signals (Proposition 1). This characterization extends results by Hermalin and Katz (1991), Feltham and Xie (1994) and Corts (2007). Interestingly, details of either the signal distribution or the agents' preferences, such as their degree of risk-aversion, are irrelevant.

Second, activities are identified if and only if each agent faces at least as many independent signals as tasks (Proposition 2). This simple characterization implies that re-organization can overcome restrictions on implementation whenever the underlying identification problem is caused by agents who decide on more than one dimension of the activity (Proposition 3). In other words, the identification problem and with it the restricted implementability must be caused by multi-tasking. As will become clear, this type of multi-tasking problem is very different from that described by Holmström and Milgrom (1991) in their classical article. The paper also provides simple conditions for when multi-tasking causes identification problems (Proposition 4, 5 and 6).

The next section introduces the class of moral hazard (or hidden action) models to which results apply and formally defines implementability and identification. Section 2 then presents the main result, namely that any activity can be implemented if and only if it is identified. Section 3 provides conditions when organization designs can overcome identification problems and hence restricted implementability. Section 4 applies the findings to established and new moral-hazard models. Section 5 explains how the findings extend and complement the existing literature and Section 6 concludes.

1 Framework

This section introduces a general analytical framework and definitions to discuss implementability and identification.

Activity and partition. Take a generic non-contractible activity that should be implemented for some exogenous reason. The activity involves n tasks $i \in N := \{1, \ldots, n\}$ for each of which there is a choice to be made $a = (a_1, \ldots, a_n)$, where the choices are from some open and convex set $A \subseteq \mathbb{R}^n$. Let $\mathcal{P}$ be some partition (disjoint decomposition) of the set of tasks N and label the elements of $\mathcal{P}$ by $l = 1, \ldots, m$. These elements, $N^l \in \mathcal{P}$, represent the set of tasks for which the choices $\{a_i\}_{i \in N^l}$ are determined by the same agent l. Denote agent l's choice vector by a^l. Observe that tasks can always be re-labeled such that the first $|N^l|$ tasks belong to agent l. So, $a = (a^l, a^{-l})$, where a^{-l} are the choices by other agents.

Agents' utility. Given activity a, agent l incurs costs $c^l(a)$, where $c^l : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable and strictly convex. Since the meaning of agent l changes with the task partition, agents' costs are allowed to change with the task partition. Agents can be rewarded in some form. For simplicity, let rewards be monetary and denote any payments to agent l by w^l. Overall, agent l's utility is strictly increasing in payments w^l and decreasing in costs: $u^l(w^l, c^l)$, where u^l is twice continuously differentiable and jointly concave in w^l and c^l. Denote agent l's outside option by $\underline{u}^l$. Since any desired marginal effect of the activity on the utility can be modeled by adjusting c^l, it is without loss of generality to assume that the marginal effect of costs on the utility is bounded: $0 > \frac{\partial u^l}{\partial c^l} \geq -1$.

Signal structure. The signal structure $\mathcal{S}$ encompasses three objects: a vector-valued function, $S = (S_1, \ldots, S_k)$, that represents k signals, a vector-valued function, $\mu = (\mu_1, \ldots, \mu_k)$, that reflects the activity's effect on signals, and continuously distributed random variables $\epsilon = (\epsilon_1, \ldots, \epsilon_k)$ with cumulative distribution function, F_ϵ, that describe influences beyond agents' control. Signal S_j is a (measurable) function of the realizations e_j of ϵ_j and the real-valued parameter, μ_j :

$$\begin{aligned} S_j : \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (\mu_j, e_j) &\mapsto S_j(\mu_j, e_j). \end{aligned}$$

$S_j(\mu_j, \epsilon_j)$ is a random variable with realizations $s_j = S_j(\mu_j, e_j)$; this random variable is assumed to be almost everywhere twice continuously differentiable and have constant support in μ_j. The parameter μ_j itself is a twice continuously differentiable and concave function of the activity a:

$$\begin{aligned} \mu_j : \mathbb{R}^n &\rightarrow \mathbb{R} \\ a &\mapsto \mu_j(a). \end{aligned}$$

Normalize signals to stochastically increase in μ_j, i.e., $S_j(\mu_j, \epsilon_j)$ first-order stochastically dominates $S_j(\mu_j', \epsilon_j)$ for $\mu_j > \mu_j'$. This is without loss of generality because any desired effect of activities on the signal can be modeled using the relationship between a and μ_j. In particular, μ_j could fall in a_i. In addition, assume that $S_j(\mu_j, e_j)$ is concave in μ_j for almost all e_j.[3]

Incentives. Agents' utility and the signal structure are assumed to be common knowledge. So, the mechanism designer can influence agents' activity by tying rewards to signals—either formally or in a self-enforcing manner. For simplicity, assume that signals are verifiable and w^l is a real-valued (measurable) function $w^l : \mathbb{R}^k \rightarrow \mathbb{R}$, that assigns a payment to agent l, $w^l(s)$, for each realization $s = (s_1, \ldots, s_k)$ of the random vector

[3]Jewitt (1988) shows that requiring concavity is milder than the assumption that the cumulative distribution function be convex in the parameter, which is typically evoked in moral-hazard models (see e.g. Rogerson 1985). In a recent contribution, Conlon (2009) discusses and extends Jewitt's results.

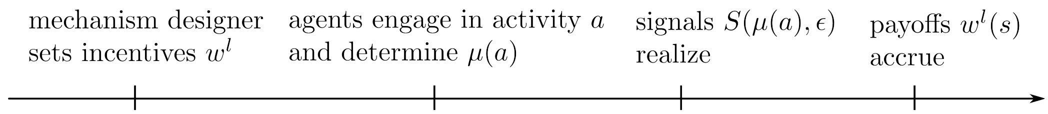

Figure 1: Sequence of Events.

$S = (S_1, \ldots, S_k)$. A specific case are payments that are *linear* in realizations: $w^l(s) = w_0^l + w_1^l s_1 + \ldots + w_k^l s_k$.

Separability of activities from noise and payments. A restriction imposed by this framework is that agent l's utility as well as the signal do not directly depend on the activity vector, a, but indirectly via the one-dimensional cost function, c^l, and parameter functions, μ_j, respectively. This precludes that components of the activity vector interact individually with payments in the utility function, e.g., $u^l(w,a) = \frac{\sqrt{w}}{a_1} - a_2$, and individual interactions with the noise term, e.g., $S_1(a, \epsilon_1) = a_1 + a_2 \cdot \epsilon_1$. However, it does not rule out that marginal effects of activities are random, e.g., $S_1(\mu_1(a), \epsilon_1) = \mu_1(a) \cdot \epsilon_1$. The separability allows us later to determine the marginal effect on agents' expected utility with the chain rule. Apart from this restriction, which is met by utility and signal functions in most moral-hazard models,[4] the setting is relatively flexible. For example, it allows for costs to change with the partition, for discretely distributed signals, and for the agents' choices to be simultaneous or sequential —this generality proves useful for applications, later.

Figure 1 gives an overview of the main elements of the model and the sequence of events. Once incentives are determined, agents $l = 1, \ldots, m$ play a game[5] and activities are implementable if they form a Nash-equilibrium of this game.

Definition 1 (Implementability)**.** *Given a signal structure $\mathcal{S}$ and a partition $\mathcal{P}$, an activity choice $a = (a_1, \ldots, a_n)$ is implementable if there are payments $\{w^l\}_{l=1,\ldots,m}$ such that the activity is a Nash equilibrium.*

Although inducing an activity is not an inference problem, one can ask whether it is possible to recover the activity from the signal distribution. Since the parameter vector, $\mu(a)$, uniquely determines this distribution, parameters can be inferred from sufficiently many (possibly infinitely many)

[4]The only exception I am aware of is the specific knowledge model by Raith (2008).

[5]Whether moves in this game are sequential or simultanous is immaterial for the later argument.

signal realizations. The question then boils down to whether activity choices a can be deduced from parameters $\mu(a)$.

Definition 2 (Identification problem)**.** *Given partition $\mathcal{P}$ and holding constant the activity of other agents, the activity of agent l is not identified if two distinct choices by this agent lead to the same signal distribution:*

$$a^l \neq \tilde{a}^l \text{ but } \mu(a^l, a^{-l}) = \mu(\tilde{a}^l, a^{-l}) \text{ for some } a^{-l}.$$

There is an identification problem if some agent's activity is not identified.

Identification is defined entirely in terms of the parameter vector, μ; the stochastic part plays no role. In particular, identification is not affected by the interdependence of errors, e.g., the variance-covariance matrix of the error vector ϵ. Moreover, identification imposes no restrictions on the interaction between different tasks—neither in the agents' cost function nor in the production of signals. All that matters is that there is an injective mapping from each agent's decision a^l to μ conditional on a^{-l}.

The next section examines the link between implementability and identification.

2 Implementability and Identification

The central aim of this section is to prove that implementability is limited if and only if there is an identification problem. For this proof, it is helpful to first characterize which activities are implementable (Proposition 1). Given this characterization, limited implementability can be traced back to the rank of the marginal effect matrix $D_{a^l}\mu(a)$ (Corollary 2). On the other hand, identification problems are also related to this rank (Proposition 2). Combining both results yields that implementability is restricted if and only if there is an identification problem.

Given some partition, signal structure and incentives, each agent l chooses activity a^l such that it maximizes his expected utility given a^{-l} :

$$a^l \in \operatorname{argmax}_{\tilde{a}^l \in A} \mathrm{E}_\epsilon \left[u^l \left(w^l \left(S \left(\mu(\tilde{a}^l, a^{-l}), \epsilon \right) \right), c^l(\tilde{a}^l, a^{-l}) \right) \right]. \qquad (1)$$

For an inner maximizer to this problem, the first-order conditions must hold. Using the separability assumption and the chain rule, the marginal effect of activities on the expected gains can be decomposed into two factors: the parameter's marginal effect on the expected value and that of the activity on the parameter. The same holds for the activity's marginal effect on costs, so that first-order conditions become:

$$\begin{aligned} & D_\mu \mathrm{E}_\epsilon \left[u^l(w^l(S(\mu,\epsilon)), c^l) \right] \cdot D_{a^l}\mu(a) \\ = \quad & -\frac{d}{dc^l}\mathrm{E}_\epsilon \left[u^l(w^l(S(\mu,\epsilon)), c^l) \right] \cdot D_{a^l}c^l(a). \end{aligned} \tag{2}$$

These first-order conditions imply a simple necessary condition. For each agent, some linear combination of his activities' marginal effect on signals must be equal to that on costs—see Lemma 1 in Appendix B. It is noteworthy that this condition is independent of agents' utility functions and hence does not depend on their degree of risk aversion, the interaction between payments and activity costs, etc. More interestingly, the condition is not only necessary but also sufficient.

Proposition 1 (Implementable activities)**.** *Given signal structure $\mathcal{S}$ and partition $\mathcal{P}$, activity $\mathring{a}$ can be implemented if and only if for each agent, some linear combination of his activities' marginal effect on signals equal that on costs given $\mathring{a}$:*

$$\lambda^l \left. D_{a^l}\mu(a)\right|_{a=\mathring{a}} = \left. D_{a^l}c^l(a)\right|_{a=\mathring{a}} \text{ with some } \lambda^l \in \mathbb{R}^k \text{ for all } l. \tag{3}$$

The proof uses linear payments: $w^l(s) = w^l_0 + w^l_1 s_1 + \ldots + w^l_k s_k$ and can be found with the proofs for all other results in Appendix A. The intuition is that the marginal gains to each agent can be set equal to marginal costs using the wage rates w^l_j. The difficulty in the proof is that adjusting w^l_j may affect agent l's marginal utility because he becomes richer (or poorer). This can be avoided by selecting a base wage w_0 dependent on w^l_j such that the agent's participation constraint binds. Then, his expected utility is constant and his marginal utility from payments is bounded away from zero. Repeating this exercise for all agents, yields that their first order conditions are only met at the desired activity. By showing that agents' utility is concave when payments are linear, the first-order conditions are not only necessary but also sufficient, and the activity indeed maximizes agents' expected utility.

Proposition 1 offers a general characterization of implementable activities for a given task partition amongst several agents. It emphasizes the fundamentals that matter for implementability: the marginal effects of activities on signals, $D_{a^l}\mu(a)$, and costs, $D_{a^l}c^l(a)$. Apart from the assumption that the agents' problems be concave, very little restrictions are placed on $\mu(a)$ and $c^l(a)$. In particular, it does not matter whether tasks are conflicting in signal generation or whether marginal costs of one task increase in the choice at another task, i.e., $\frac{\partial^2 c^l(a)}{\partial a_i \partial a_{i'}} > 0$. Since the proof uses linear payments to show that condition (3) suffices for implementation, we immediately obtain the following corollary.

Corollary 1. *Any activity $\mathring{a}$ that can be implemented can also be implemented with linear payments.*

The corollary shows that imposing linearity when trying to find out which activities are implementable is without loss of generality. It thus justifies the approach by Feltham and Xie (1994) and Corts (2007) of restricting attention to linear payments when examining implementation.

Proposition 1 is also helpful from a technical point of view because it reduces the question of implementability to that of the existence of a solution to a linear equation system. This means that standard results from linear algebra can be used to check for implementability. For example, the result that a linear equation system has a solution if the rank of the extended coefficient matrix is equal to the rank of the coefficient matrix itself yields the following corollary.

Corollary 2 (Rank characterization for implementable activities)**.** *Given signal structure $\mathcal{S}$ and partition $\mathcal{P}$, activity $\mathring{a}$ is implementable if and only if for all agents, marginal costs are in the row-space of the matrix describing their marginal effects on signals:*

$$rank\,(D_{a^l}\mu(a))\Big|_{a=\mathring{a}} = rank \begin{pmatrix} D_{a^l}\mu(a) \\ D_{a^l}c^l(a) \end{pmatrix}\Big|_{a=\mathring{a}} \text{ for all } l. \tag{4}$$

This characterization of implementable activities is useful because it expresses implementability using the rank of the marginal effect matrix, $D_{a^l}\mu(a)$. As the next proposition shows, the same rank also matters for identification.

Proposition 2 (Identification and signal independence)**.** *Given signal structure $\mathcal{S}$ and partition $\mathcal{P}$, agents' activity is not identified if and only if some agent l faces less independent signals than tasks:*

$$rank\,(D_{a^l}\mu(a)) < |N^l| \text{ for some } l \text{ and } a.$$

The intuition for the result is the following. Suppose there are less independent signals than tasks for some agent or more formally that the rank of the marginal effect matrix is not large enough for some choice a. Then, there is a direction in which a change of the agent's decision does not affect the parameter vector. Choices along this direction can hence not be identified and the mechanism designer faces an identification problem. Conversely, having sufficiently many independent signals, i.e., a sufficiently large rank of the marginal effect matrix, means that any direction leads to a different parameter vector, so that the agent's activity is identified.

The rank condition is somewhat weaker than invertibility of the agents' marginal effect matrix. Invertibility would in addition require that the matrix is quadratic, i.e., the number of signals would have to be the same as the number of tasks. The proposition can also be re-interpreted as a characterization of organization designs that are free from identification problems. In the next section, this interpretation is used to derive necessary and sufficient conditions for organization designs to solve identification problems.

Using that implementation and identification depend on the rank of the same matrix yields the following central result.

Theorem 1 (Implementability and identification)**.** *Given signal structure $\mathcal{S}$ and partition $\mathcal{P}$, implementability is limited if and only if there is an identification problem.*

The logic of the proof is the following. Whenever there is no identification problem, there are at least as many independent signals as tasks for all agents and activities by Proposition 2. The central step in the proof is to show that this is fulfilled whenever the rank condition (4) in Corollary 2 holds.

The proof is indirectly based on Proposition 1, which uses linear payments. Linear payments are problematic in the canonical hidden-action model with limited liability and a risk-neutral agent, i.e., $u^l(w^l, c^l) = w^l - c^l$ with the additional restriction $w^l \geq 0$, because linearity implies that the agent may lose money for some (possibly rare) signal realizations.[6] Still, identification also leads to implementability in this setting. Since activities are identified, there are signal realizations which are most likely when each agent selects the desired choice; tying a finite bonus to these realizations then induces the desired activity (see Proposition 7 in Appendix C).

The theorem pins down the signal imperfection that hampers implementability: the lack of identification. Interestingly, only the signal structure matters for implementability and not, for example, agents' preferences. Moreover, it is only the systematic effect of the activity on signals that matters and not the distribution of the error terms.

[6]For a textbook treatment see, for example, Macho-Stadler and Perez-Castrillo (1997).

The theorem suggests looking at limited implementability through the lens of an inference problem. While this analogy is helpful, it is not entirely correct. If the activity is identified according to Definition 2, an econometrician needs a sufficient number of independent realizations s of the signal vector S to deduce activity a^l from signals (possibly infinitely many). Moreover, the econometrician has to know the choices of the other agents a^{-l}. On the other hand, knowing choices a^{-l} is not required and only one realization of S suffices in order to induce the desired activity. The reason is that conditioning payments on this realization provides the correct marginal gains and hence ex-ante incentives for agents to engage in this activity, i.e., to choose the desired $\mathring{a}$. Given these incentives, the mechanism designer knows that agents choose correctly and there is no need to deduce choices from signals.

3 Identification by Organization Design

The previous section has shown that implementability is limited whenever there are identification problems. Moreover, Proposition 2 introduced a simple characterization for organization designs that are free of such problems. Building on this proposition, the present section provides general conditions under which organization design can overcome identification problems (Proposition 3 and Proposition 4) and more specific conditions on the number of agents, independent signals, and tasks (Proposition 5 and Proposition 6). The next proposition links multi-tasking to the question whether organization design can solve identification problems.

Proposition 3 (Identification problems, organization design, and multi–tasking)**.** *Given signal structure $\mathcal{S}$, identification problems can be solved by some organization design $\mathcal{P}$ if and only if they are caused by an agent who carries out more than one task (multi-tasking).*

The finest partition $\{\{1\}, \ldots, \{n\}\}$, i.e., assigning a different agent to each task, is an example of an organization design. If the problem is caused by multi-tasking, this partition solves the problem. Conversely, an identification problem that can be solved by some partition can also be alleviated with the finest partition. The intuitive reason is that the finest partition leaves agents less scope to generate signals than any other partition (the formal proof is with all other proofs in Appendix A). With Proposition 3, it becomes simple to determine whether identification by organization design is possible: we only need to check whether the finest partition can solve the identification problem. This insight allows us to characterize when organization design helps with identification.

Proposition 4 (Identification by organization design)**.** *Given signal structure $\mathcal{S}$, some organization design $\mathcal{P}$ solves any identification problems if and only if different choices at some task i lead to different signal distributions (holding constant the choices at all other tasks):*

$$\textit{for all } i \textit{ and } a_i \neq \tilde{a}_i : \quad \mu(a_1, \ldots, a_i, \ldots, a_n) \neq \mu(a_1, \ldots, \tilde{a}_i, \ldots, a_n).$$

For the proof, recall that by Proposition 3, it suffices to examine whether the finest partition solves an identification problem in order to know whether identification by organization design is possible. Given that each task is carried out by a different agent, identification boils down to the condition that the parameter is affected by each a_i when other choices remain the same. Re-phrasing the proposition, identification by organization design works if the parameter vector is an injective function of each decision a_i for given constant choices at all other tasks. Notice that conditional injectivity is sufficient. In particular, μ does not have to be injective in a. It may thus be impossible to infer a from signals and still organization design overcomes the identification problem–as seen in the program stability example in the introduction.

According to Proposition 3, organization design can only overcome identification problems if they are caused by multi-tasking. Also, we know from Proposition 2 that entirely eliminating multi-tasking is not necessary to overcome identification problems: an agent may well carry out more than one task if there are sufficiently many independent signals. The rest of this section provides necessary and sufficient conditions on the number of workers, independent signals and tasks such that identification problems can be alleviated.

Proposition 5 (Necessary condition)**.** *Given signal structure $\mathcal{S}$, identification problems can only be alleviated by some partition $\mathcal{P}$ if the total number of independent signals, $\tilde{k}$, exceeds the average number of tasks per agent:*

$$\tilde{k} \geq \frac{n}{m},$$

where $\tilde{k} := \min_a \mathit{rank}(D_a \mu(a))$.

The result is based on Proposition 2. The number of independent signals for some agent l can at most be the overall number of independent signals,

$\tilde{k}$. If an agent is assigned more tasks than the total number of independent signals, this will lead to an identification problem. The maximal number of tasks that can possibly be assigned to each agent is thus $\tilde{k}$. With m agents, the maximal possible number of tasks that may still lead to identification (and implementability) is hence $m \cdot \tilde{k}$.

While the condition in the proposition is necessary, it is not sufficient. As a counter-example, consider a situation in which four tasks affect two signals but only the second reflects the fourth task: $S_j(\mu_j, \epsilon_j) = \mu_j + \epsilon_j$ for $j = 1, 2$ with $\mu_1 = a_1 + a_2 + a_3$ and $\mu_2 = a_1 + a_2 + a_3 + a_4$. Then, the marginal effect matrix is:

$$D_a(\mu(a)) = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

The minimum rank of the matrix is two because the two rows are linearly independent. With two agents, the average number of tasks per agent is likewise two and the necessary condition in Proposition 5 is met: $\tilde{k} = 2 = 4/2 = n/m$. Still, the identification problem cannot be eliminated by any partition among the two agents. Consider the first three tasks $\{1, 2, 3\}$. At least two of these tasks have to be carried out by the same agent l. For this agent, $\text{rank}(D_{a^l}\mu(a)) = 1$ while he carries out two tasks. By Proposition 2, there is thus an identification problem and the condition that the average number of tasks per agent must be larger than the total number of independent signals is necessary but not sufficient. A sufficient condition is given in the next result.

Proposition 6 (Sufficient condition)**.** *Given signal structure $\mathcal{S}$ and an identification problem that is caused by multi-tasking, the identification problem can be alleviated by some partition $\mathcal{P}$ if the number of agents and independent signals strictly exceeds the number of tasks:*

$$m + \tilde{k} > n,$$

where $\tilde{k} := \min_a \textit{rank}(D_a\mu(a))$.

The intuition behind this result is the following. With a total number of $\tilde{k}$ independent signals, $\tilde{k}$ tasks can be carried out by one agent without causing an identification problem. To guarantee that the remaining $n - \tilde{k}$ tasks are not resulting in such a problem, they have to be assigned to $n - \tilde{k}$ agents. Thus, $n - \tilde{k} + 1$ agents are sufficient to avoid the identification problem.

The condition, $m + \tilde{k} > n$, is sufficient but not necessary. To see this, suppose again that there are four tasks but that each task only affects one of two signals: $S_j(\mu_j, \epsilon_j) = \mu_j + \epsilon_j$ with $\mu_1(a) = a_1 + a_3$ and $\mu_2(a) = a_2 + a_4$ The marginal effect matrix is:

$$D_a(\mu(a)) = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix},$$

so that $k = \tilde{k} = 2$. At the same time, there are two agents, $m = 2$, so that $m + \tilde{k} = 4$, which is not strictly larger than the number of tasks, $n = 4$. However with partition $\{\{1,2\},\{3,4\}\}$, each agent has two independent signals and activities are identified because each agent carries out only two tasks.

4 Applications

This section exploits the generality of the framework and applies the results to established and new moral-hazard models in order to derive insights into identification and implementability in these models.

4.1 Advocates, Specialists, and Joint Accountability

Three contributions that particularly emphasize the beneficial effects of organization design on implementation are the advocates model (Dewatripont and Tirole, 1999), the specialization model (Ratto and Schnedler, 2008),[7] and the accountability model (Corts, 2007). These models differ in various dimensions (e.g., number and character of tasks, number of signals, flexibility of number of agents). Despite these differences, results are driven by the same fundamental principle: identification by organization design.

All three contributions start out with an organization structure that is plagued by an identification problem (single investigator, single worker, individual accountability) and limited implementability. Moreover, signals are injective in activity choices at all tasks, so that these identification problems are caused by multi-tasking (by Proposition 4) and can be overcome by organization design (Proposition 3). The proposed alternative organization designs (advocacy, specialization, joint accountability) exploit this opportunity by increasing the number of independent signals relative to the number of tasks, so that activity choices become identified (by Proposition 2) and implementability is no longer limited (by Theorem 1).

[7] Inderst and Ottaviani (2009) discuss an application similar to Ratto and Schnedler (2008) in which specialization overcomes restricted implementability.

In all three contributions, identification involves moving from a situation in which each signal is determined by only one agent to one in which multiple agents affect the same signal. This suggests that identification comes at the price of joint instead of individual accountability.[8] This, however, is not the case. Even if agents are already jointly accountable to begin with, reorganization may lead to identification.

In order to see this, consider the following variation of the program stability example. Suppose that in addition to programming effort, a_1, and debugging effort, a_2, there is also sales effort, a_3. The stability of the program, μ_1, which positively depends on the first two efforts, can be measured in form of some signal, say, a test run S_1 that may be successful or not. In addition, the client's inclination to buy the product, μ_2, increases with the program's stability μ_1 as well as sales effort and leads to a sales signal, S_2, which reflects whether the client buys the program or not.

First, take the following task assignment: a production agent is responsible for (a_1, a_2) and a sales agent decides on a_3. With this assignment, both agents are jointly responsible for the sales signal, i.e., they both determine the distribution of signals.[9]

Now consider a different partition of tasks, where one agent is responsible for programming only, while the other chooses debugging and sales effort (a_2, a_3). Both agents affect both signals and are hence jointly responsible. On the other hand, the new partition exhibits no identification problem. Given programming effort a_1, the signal distributions, parameterized by (μ_1, μ_2), reflect any change in debugging and sales effort and vice versa. Thus changing from one partition to the other yields identification and unrestricted implementation—although the former does not involve individual accountability. The introduction of joint accountability is hence not a necessary condition for identification.

4.2 Complementarities versus Control

Although fine partitions of tasks facilitate identification and thus give managers better control, real production typically involves task bundling. Per-

[8] While joint accountability has a smack of a free-rider problem, such problems are absent from all three models because the principal acts as a budget breaker–see Holmström (1982).

[9] Enriching the model with respective assumptions on risk-aversion and signals, both agents are optimally paid on the basis of both signals given this task assignment and hence their pay depends on the other agents' activity.

haps the main reason for task bundling are complementarities, either for technical reasons or because agents like variety.[10] The gains from complementarities then need to be traded off against the gains from control. Arguably, optimal organization design results from this fundamental trade-off and is hence driven by two parameters: (i) the degree of complementarity and (ii) the importance of influencing activity choices. In the following, this argument is formalized in a simple model, which is another variation of the (two-task) program stability example from the introduction.

Suppose programming and debugging exhibit complementarities and that the gains of complementarities only manifest if both tasks are carried out by the same agent. Let costs amount to $c^B(a) = \frac{a_1^2 + a_2^2 - \gamma a_1 a_2}{2}$ if tasks are bundled and to $c^j(a) = \frac{a_j^2}{2}$ if each task $j = 1, 2$ is carried out by a different agent j, so that $\gamma \in [0, 2)$ describes the extent of complementarities. Standardize the agents' outside option to zero. The only observable variable is the program's stability, μ. For simplicity, let μ be linear in both efforts: $\mu(a) = a_1 + a_2$. In order to model the importance of control, assume that the principal's utility is not only determined by stability but that she has a larger benefit from programming: $u^P(a) = \mu(a) + \delta a_1 - w$, where w stands for any wages paid to agents and $\delta \geq 0$ captures the degree to which control matters to the principal.

The following analysis proceeds in three steps. First, the optimal contract and maximal surplus under task bundling is determined. Second, the same is done for task separation. Finally, surpluses are compared to show that bundling is optimal whenever the principal wants to influence the agent's decision (large δ) and gains from complementarity are limited (small γ). In the analysis, attention is restricted to linear contracts. This is without loss of generality because the only friction in the model is that some activities may not be implementable and linear contracts suffice to overcome this friction (see Corollary 1).

Task bundling leads to an identification problem because the first agent faces two tasks but only one signal (by Proposition 2) and implementability is restricted (by Theorem 1). The principal's program is:

$$\begin{aligned}
&\max_{w_0, w_1} && (1+\delta)a_1 + a_2 - w_0 - w_1(a_1 + a_2) \\
&\text{such that} && (a_1, a_2) \in \arg\max_{\tilde{a}} w_0 + w_1(\tilde{a}_1 + \tilde{a}_2) - c^B(\tilde{a}) && \text{(IC)} \\
&\text{and} && w_0 + w_1(a_1 + a_2) - c^B(a) = 0. && \text{(PC)}
\end{aligned}$$

[10] Other reasons could be insurance or hiring costs (see e.g. Corts, 2007).

The incentive constraint (IC) boils down to the implementability condition from Proposition 1:

$$w_1 D_{(a_1,a_2)}\mu(a) = D_{(a_1,a_2)}c^B(a) \;\text{ or }\; w_1(1,1) = (a_1 - \frac{\gamma}{2}a_2, a_2 - \frac{\gamma}{2}a_1).$$

Eliminating w_1 in this equation system yields that only identical choices on both tasks are implementable, i.e., $a_1 = a_2$. Plugging this condition as well as the participation constraint (PC) into the maximization program yields

$$\max_{a_1} \quad (1+\delta)a_1 + a_1 - \frac{a_1^2 + a_1^2 - \gamma a_1 a_1}{2}. \tag{5}$$

From this, the optimal activity choices under task bundling can be computed to be $a_1^{\mathrm{B}} = a_2^{\mathrm{B}} = \frac{2+\delta}{2-\gamma}$. The respective surplus is

$$\pi^{\mathrm{B}} := \frac{1}{2}\frac{(2+\delta)^2}{(2-\gamma)}.$$

Task separation leads to conditionally identified activities because each agent faces as many independent signals as tasks (by Proposition 2) and implementability is unrestricted (by Theorem 1). Plugging in the participation constraints, we get the following maximization program:

$$\max_{a_1,a_2} \quad (1+\delta)a_1 + a_2 - \frac{a_1^2}{2} - \frac{a_2^2}{2} \tag{6}$$

This program leads to the optimal activity choices: $a_1^{\mathrm{S}} = (1+\delta)$ and $a_2^{\mathrm{S}} = 1$, which yields a surplus of

$$\pi^{\mathrm{S}} := \frac{(1+\delta)^2}{2} + \frac{1}{2}.$$

The analysis implies that task separation is optimal whenever

$$\pi^{\mathrm{S}} > \pi^{\mathrm{B}} \;\text{ or }\; \frac{(1+\delta)^2 + 1}{(2+\delta)^2} > \frac{1}{2-\gamma}.$$

The left-hand side increases in δ and the right-hand side in γ. So, separating tasks becomes relatively more attractive if control is important and complementarities are small.

4.3 Implementability in Single-Task Models

In the pioneering works on the moral-hazard model by Holmström (1979, 1982) and Shavell (1979) and many ensuing contributions, the agent faces only one decision: how much effort, $a \in \mathbb{R}$, to exert. The focus of this literature is not on implementability but on the trade-off between incentives and insurance. The results from the present paper provide an explanation why this might be the case: in these models, any activity can be implemented.[11]

The intuition as to why any activity choice can be implemented is the following. More effort is assumed to stochastically increase output Y in the sense of first-order stochastic dominance: $G_a(y) > G_{\tilde{a}}(y)$ for $\tilde{a} > a$, where $G_a(y)$ is the cumulative distribution function of Y given a. This so-called Mirrlees representation can to be transformed to the state-space representation used here[12] by setting $\mu(a) = a$ for all a and an appropriate choice of S_j and ϵ_j (see Lemma 6 in Appendix C). The parameter μ is thus injective in the one-dimensional activity, so that the activity is identified (by Propositions 2 and 4) and implementation is not limited (by Theorem 1).

4.4 Multi-Tasking Problems

Corts (2007) claims that in large parts of Holmström and Milgrom's seminal paper on multi-tasking (1991) "there is no multi-task problem." This section clarifies, supports and extends Corts' claim: Holmström and Milgrom's article features no identification problem that is caused by the fact that one agent carries out more than one task. This observation seems to contradict that Holmström and Milgrom's article is known for having provided the first formal analysis of multi-tasking problems. The seeming contradiction results from the fact that the term 'multi-tasking problem' refers to two very different phenomena. On the one hand, multi-tasking can negatively affect the trade-off between insurance and incentives,[13] which is the problem described by Holmström and Milgrom. On the other hand, we have seen here that multi-tasking can lead to identification problems and limit implementability, which is probably the multi-tasking problem referred to by Corts (2007).

The following substantiates the claim that multi-tasking causes no re-

[11] See Section 4.6 for an example of a single-task model with restricted implementability.

[12] In using the state-space representation, I follow the tradition of Spence and Zeckhauser (1971), Ross (1973), Jewitt (1988), and Conlon (2009)

[13] See Prendergast (1999, 2002) for a critical empirical overview on the evidence concerning this trade-off.

strictions on implementability in any of the model variations discussed in Holmström and Milgrom's article. These variations employ either what Corts (2007) calls the *one-signal-per-task assumption* or what may be called a *no-signal-for-some-task assumption*.[14] The first assumption describes that for each task i, there is a signal j that strictly increases in the agent's choice at that task and is unaffected by other choices: $\frac{\partial \mu_j}{\partial a_i} > 0$ if $i = j$ and $\frac{\partial \mu_j}{\partial a_i} = 0$ otherwise. This assumption implies that there are as many independent signals as tasks: $\text{rank} D_a \mu = n$. So even if a single agent is responsible for all n tasks, this agent's activity is identified (by Proposition 2) and any activity choice can be implemented (by Theorem 1). The second assumption means that some task i affects no signal j: $\frac{\partial \mu_j}{\partial a_i} = 0$ for all j. This assumption clearly leads to an identification problem. However, this problem cannot be alleviated by eliminating multi-tasking or any other form of re-organization (by Proposition 4).

The models discussed in Holmström and Milgrom (1991) thus concern two extreme situations: those in which *any* and those in which *no* assignment of tasks leads to identification. Examining only these extremes, identification by organization design cannot be studied. Accordingly, Holmström and Milgrom's results on optimal job design are not driven by the desire to induce new activities.

4.5 Multi-Tasking with Linear Signals

Holmström and Milgrom (1991) inspired a burgeoning literature in accounting (see e.g. Feltham and Xie, 1994; Datar, Cohen Kulp, and Lambert, 2001) and some contributions in labor economics (Baker, 2000, 2002; Schnedler, 2008), which examine multi-tasking in a particularly tractable framework. Among other things, this framework assumes that signals are linear in activity choices. The literature starts with the observation that the principal's power to implement activities is restricted and proceeds to examine how congruency between signals and the principal's benefit affect optimal (linear) contracts. It neither identifies the source of restricted implementability nor does it consider whether and when organization design may solve the problem.

With the tools introduced here, this gap can be closed. Given linear

[14] For example, Holmström and Milgrom's job design model uses the one-signal-per-task assumption, while their home contractor model or asset enhancement models employ the no-signal-for-some-task assumption.

signals, organization design can solve identification problems unless there is no signal for some task. The reason is the following. Since signals are linear in activities, $\mu_j(a) = \mu_{j1}a_1 + \ldots + \mu_{jn}a_n$, the parameter vector μ is injective in the activity choice at each task i as long as there is some signal j for each task i with $\mu_{ji} \neq 0$. If such signals exist, the identification problem is hence due to multi-tasking (Proposition 3) and can be solved by organization design (Proposition 4).

4.6 Identification with Biased Reports

In many circumstances, incentives are based on subjective evaluations or reports. Experts writing these reports rarely evaluate behavior in exactly that way which is required to induce the desired activity choice; in short, reports are biased. Under these circumstances, a second report can be valuable even if it is more biased.

In order to support this claim, consider a risk-neutral agent whose activity choice, $a \in \mathbb{R}$, affects the probability of report j being favorable ($S_j = 1$) or not ($S_j = 0$) with $j = 1, 2$. More specifically, let the probability of a favorable report, $P(S_j = 1|\mu_j)$, decrease in the distance μ_j between the activity a and some benchmark $\alpha_j : \mu_j(a) = (a - \alpha_j)^2$. For simplicity, let the principal's benefit from the activity be the activity itself and the agent's cost be quadratic. Then, the joint surplus amounts to: $a - c(a) = a - \frac{a^2}{2}$ and is maximized at $a = 1$. In order to reflect that both reports are biased and the second even more so, suppose that the two benchmarks are too low: $0 < \alpha_2 < \alpha_1 < 1$. Which report is used in optimal contracts and when is a second report valuable?

Each of the two signals alone is not injective in the activity choice. For example, $\mathring{a} := \alpha_j + 1$ and $\tilde{a} := \alpha_j - 1$ both lead to $\mu_j = 1$. With only one signal, the agent's activity is thus not identified and implementation is limited. Moreover, the problem is not due to multi-tasking and organization design offers no viable identification strategy. Activity choice $\mathring{a}$ can only be induced if some linear combination of the marginal effect on the signal, λ, equals marginal costs: $\lambda(-2(\mathring{a} - \alpha_j)) = \mathring{a}$ (by Proposition 1). Solving for $\mathring{a}$ yields: $\mathring{a}(\lambda) = \frac{\alpha_j}{\frac{1}{2\lambda}+1}$. While the activity choice increases in λ, it is bounded: $\lim_{\lambda\to\infty} \mathring{a}(\lambda_1) = \alpha_j$. Factoring in the agent's participation constraint, the principal's net gain equals the joint surplus: $a - \frac{a^2}{2}$. This net gain strictly increases in a for $a < \alpha_j < 1$ and thus attains its supremum at $a = \alpha_j$. If the principal had to select one report, she would take the less biased one and obtain a surplus that is arbitrarily close to

$$\pi^{1\mathrm{R}} := \alpha_1 - \frac{\alpha_1^2}{2}.$$

With two reports, the identification problem is solved. In order to see this, consider the marginal effect matrix:

$$D_a\mu = \begin{pmatrix} -2(a-\alpha_1) \\ -2(a-\alpha_2) \end{pmatrix}.$$

This matrix has a minimal rank of one because one of the two entries is always different from zero. Hence, a single agent with one task faces one independent signal and his activity is identified (by Proposition 2). Any activity can thus be implemented, in particular, the first-best choice, $a = 1$. The surplus with both reports thus amounts to:

$$\pi^{2\mathrm{R}} := \frac{1}{2}.$$

Accordingly, the additional report has a value of at least

$$\pi^{2\mathrm{R}} - \pi^{1\mathrm{R}} = \frac{1}{2} - \alpha_1 + \frac{\alpha_1^2}{2} = \frac{1}{2}(1-\alpha_1)^2.$$

This value is independent of α_2 and strictly positive (as long as the first report is biased). The second report thus adds value irrespective of its bias and this value becomes larger, the larger the bias of the first report, i.e., the smaller α_1.

4.7 Identification and Sufficient Statistics

The notion of identification advocated here is related to that of a sufficient statistic. Holmström (1979) famously suggested to assess the value of an additional signal S_2 by checking whether the original signal S_1 is a sufficient statistic for activity a. An additional signal S_2 that is informative in the sufficient statistic sense, however, does not necessarily identify the activity.

Consider one agent who faces two tasks and two signals, $S_j(\mu_j, \epsilon_j) = \mu_j + \epsilon_j$ with $\mu_1(a) = \mu_2(a) = a_1 + a_2$ and let ϵ_j be an independently standard normally distributed error term. Using only S_1 leads to the —by now familiar— identification problem and this problem persists even if both signals are used because marginal effects of the activity on both signals are identical. On the other hand, signal S_2 conditional on S_1 is normally distributed with mean a_1+a_2 and thus depends on activities. Signal S_2 is hence informative about the activity in Holmström's sense but useless in terms of identification and implementability.

Based on his concept, Holmström (1979) formulates the well-known sufficient statistic result. This result is restricted to distributions from the exponential class with rank one (Amershi and Hughes, 1989) and can fail if activities are multi-dimensional (Holmström and Milgrom, 1991) or if income affects marginal costs of effort (Schnedler, 2010). In contrast, results on identification derived here are independent of the specific distribution, hold for single- and multi-dimensional activities, and allow for interaction between effort and wealth.

5 Related Literature

Probably the first to analyze implementability in moral-hazard models are Hermalin and Katz (1991). They observe that a specific choice is only implementable if there is no (less costly) way for the agent to produce the same signal distribution—see their Proposition 2. Theorem 1 in the present paper builds on this idea and applies it to multi-dimensional activities with continuous choice sets. In addition, it allows for the possibility of multiple agents, so that organization design can be discussed.

For multi-dimensional activities, linear performance measures (signals), normally distributed error terms, quadratic costs and either negative exponential utility or risk-neutrality, Feltham and Xie (1994) find that an activity is implementable if and only if "it is spanned by the set of performance measure coefficients"—see their Appendix B. Like Feltham and Xie (1994), Corts (2007) restricts attention to linear signals; he allows for separable cost functions and a general noise term but limits attention to linear contracts and only considers risk-neutral agents. He finds that optimal linear contracts induce first-best efforts whenever signals "span the [agent's] set of tasks" (see his Proposition 6) and observes that this is only possible if the average number of tasks per agent is at most the number of signals. At first glance, Corts' finding may not concern implementability. However, it is only implementability that prevents the first-best from being induced if agents are risk-neutral.

Feltham and Xie as well as Corts' findings are extended beyond linear signals and specific utility and cost functions in Proposition 1, which states that an activity is implementable if the vectors describing the activity's marginal effect on signals span the vector describing the activity's marginal effect on costs. Moreover, Corollary 1 provides a justification for Corts' approach to limit attention to linear contracts, while Proposition 5 tightens his necessary condition for implementing the first-best by considering only independent instead of all signals, and Proposition 6 adds a sufficient condition.

While the present paper focuses on whether specific ways can be induced to produce a given observable output, the literature on partnerships is concerned with the complementary question of inducing a certain (joint) output (see e.g. Legros and Matsushima, 1991; Strausz, 1999; Battaglini, 2006). Implementing this output is complicated because budgets must balance in partnerships. On the other hand, it is efficient if partners produce a given output in the least-costly way and the conflict between a desired way of producing a signal, say $\mathring{a}$, and agents' chosen method, say $\tilde{a}$, does not arise.

Some of the issues in the present paper, e.g., implementability or organizational structure, are also relevant for mechanism design with hidden information.[15] They are examined here from a hidden action perspective. Finally, organizational form can also be discussed outside the principal-agent paradigm—see Borland and Eichberger (1998) for an overview.

6 Conclusion and Discussion

In the first step of their famous analysis of the principal agent-problem, Grossman and Hart (1983) determine the least-costly way of implementing a specific activity. They also observe that it may sometimes be prohibitively costly to implement an activity. However, they do not examine when this is the case. The present paper addresses this question. It advocates incorperating the notion of *identification* into agency theory in addition to other originally econometric concepts such as 'likelihood ratio' (Mirrlees, 1979; Rogerson, 1985), 'sufficient statistic' (Holmström, 1979, 1982), and 'hazard rate' (see e.g. Gibbons, 1987). Identification describes the signal imperfection that restricts implementability: any activity choice can be induced if activities are

[15]A classical problem is that of implementing a social choice rule (see e.g. Fudenberg and Tirole, 1991, Chapter 7.1). Mookherjee (2006) provides an excellent survey on organizational structures in hidden information models, in particular on the question when decisions should be decentralized. Alonso and Matouschek (2008) are concerned with limiting choice sets if agents are better informed; Raith (2008) examines how to provide incentives to work to such agents.

identified (Theorem 1).

Dealing with performance measures in accounting, Hemmer (1998) suggests that the informational content of the same performance measure system can be altered when the organizational design changes. This idea also applies to identification: even if the signal structure remains the same, reorganization can lead to identification whenever a change in some dimension of an activity affects observables (Proposition 4). The characteristics of organization designs that solve the identification problems echo Tinbergen's dictum (1952) that there need to be as many instruments as goals: the number of independent signals for an agent must be at least as large as his number of tasks (Proposition 2).

Identification by organization design is driven by two crucial assumptions. First, it must be possible to enforce decision rights. For example, agents must be prevented from carrying out tasks that they are not supposed to, while agents who should work on these tasks must be able to do so. This points to one possible reason why restricted access to tools and production sites is common in many firms and organizations: it gives the management more control over how outcomes are achieved. Second, if agents collude, then identification problems cannot be solved by partitioning tasks. Identification becomes possible because partitions generate a non-cooperative game in which each agent's equilibrium strategy can be separately manipulated. If agents were colluding, which requires that they can (formally or informally) condition payments on activities although the mechanism designer cannot, they would jointly agree to chose the least-costly activity in order to produce a given outcome; no other way of producing this outcome could be implemented.

The results from the present paper uncover why certain organization designs, which are proposed in the literature, such as advocates, specialists, or joint accountability, are advantageous: these designs overcome identification problems and generate unlimited implementability. The findings can also be employed to model organization design as being driven by a fundamental trade-off between control, on the one hand, and the gains from complementarities, on the other hand. Moreover, they offer an explanation why in hidden action models (very unlike in hidden information models), implementability has received so little attention: in the early single-task models, any activity can be implemented. In addition, results help to pin down a specific type of problem caused by multi-tasking. If an agent carries out more than one task, this can create identification problems and may thus limit implementability. This type of problem is latent in many models but very different from that described by Holmström and Milgrom (1991). As shown, multi-tasking causes no identification problems in this article and hence does not restrict which activities can be implemented. Finally, the findings imply that soliciting biased reports can have a value because doing so generates identification and thus facilitates implementation.

The present paper shows that given identification, payments can be set such that any desired activity is induced as a Nash equilibrium. This leaves several questions open. Can unique implementation be achieved? Can activities be implemented in dominant strategies or as an attractor of best-response dynamics? All these questions are beyond the scope of the paper and left for future research. While the paper links implementability, identification, and organization design in a rather general framework, it only provides a first-step in the analysis of implementability in moral hazard models. The generality of the structure only permits relatively fundamental statements about existence and form of organization designs that solve identification problems. Finer predictions require additional assumptions, for example, on which tasks are separable and how activities are reflected in signals.

References

Alonso, R., and N. Matouschek (2008): "Optimal Delegation," *Review of Economic Studies*, 75(1), 259–293.

Amershi, A. H., and J. S. Hughes (1989): "Multiple Signals, Statistical Sufficiency, and Pareto Orderings of Best Agency Contracts," *RAND Journal of Economics*, 20(1), 102–112.

Baker, G. (2000): "The Use of Performance Measures in Incentive Contracting," *American Economic Review*, 90(2), 415–420.

——— (2002): "Distortion and Risk in Optimal Incentive Contracts," *Journal of Human Resources*, 37(4), 728–751, Special Issue on Designing Incentives to Promote Human Capital.

Battaglini, M. (2006): "Joint production in teams," *Journal of Economic Theory*, 130(1), 138 – 167.

Borland, J., and J. Eichberger (1998): "Organizational Form Outside the Principal-Agent Paradigm," *Bulletin of Economic Research*, 50(3), 201–27.

Conlon, J. R. (2009): "Two New Conditions Supporting the First-Order Approach to Multisignal Principal-Agent Problems," *Econometrica*, 77(1), 249–278.

Corts, K. S. (2007): "Teams versus Individual Accountability: Solving Multitask Problems Through Job Design," *RAND Journal of Economics*, 38(2), 467–479.

Datar, S., S. Cohen Kulp, and R. A. Lambert (2001): "Balancing Performance Measures," *Journal of Accounting Research*, 39(1), 75–92.

Dewatripont, M., I. Jewitt, and J. Tirole (2000): "Multitask Agency Problems - Focus and Task Clustering," *European Economic Review*, 44(4-6), 869–877.

Dewatripont, M., and J. Tirole (1999): "Advocates," *Journal of Political Economy*, 107(1), 1–39.

Feltham, G. A., and J. Xie (1994): "Performance Measure Congruity and Diversity in Multi-Task Principal/Agent Relations," *The Accounting Review*, 69(3), 429–453.

Fudenberg, D., and J. Tirole (1991): *Game Theory*. The MIT Press, Cambridge, Massachusetts.

Gibbons, R. (1987): "Piece-Rate Incentive Schemes," *Journal of Labor Economics*, 5(4), 413–429.

——— (1998): "Incentives in Organizations," *Journal of Economic Perspectives*, 12(4), 115–132.

Grossman, S. J., and O. D. Hart (1983): "An Analysis of the Principal-Agent Problem," *Econometrica*, 51(1), 7–46.

Hemmer, T. (1998): "Performance measurement systems, incentives, and the optimal allocation of responsibilities," *Journal of Accounting and Economics*, 25(3), 321–347.

Hermalin, B., and M. Katz (1991): "Moral Hazard and Verifiability: The Effects of Renegotitation in Agency," *Econometrica*, 59(6), 1735–1753.

Holmström, B. (1979): "Moral Hazard and Observability," *Bell Journal of Economics*, 10(1), 74–91.

——— (1982): "Moral Hazard in Teams," *Bell Journal of Economics*, 13(2), 324–340.

Holmström, B., and P. Milgrom (1991): "Multitask Principal-Agent-Analysis: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization*, 7, 24–52, special issue.

Inderst, R., and M. Ottaviani (2009): "Misselling Through Agents," *American Economic Review*, 99, 883–908.

Jewitt, I. (1988): "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, 56(5), 1177–1190.

Legros, P., and H. Matsushima (1991): "Efficiency in partnerships," *Journal of Economic Theory*, 55(2), 296–322.

Macho-Stadler, I., and D. Perez-Castrillo (1997): *An Introduction to the Economics of Information: Incentives and Contracts*. Oxford University Press, Oxford.

Mirrlees, J. A. (1979): "The Implications of Moral Hazard for Optimal Insurance," Seminar given at converence held in honor of Karl Borch, Bergen, Norway.

Mookherjee, D. (2006): "Decentralization, Hierarchies, and Incentives: A Mechanism Design Perspective," *Journal of Economic Literature*, 44(2), 367–390.

Prendergast, C. (1999): "The Provision of Incentives in Firms," *Journal of Economic Literature*, XXXII, 7–63.

——— (2002): "The Teneous Trade-Off Between Risk and Incentives," *Journal of Political Economy*, 110(5), 1071–1102.

Raith, M. (2008): "Specific Knowledge and Performance Measurement," *The RAND Journal of Economics*, 39(4), 1059–1079.

Ratto, M., and W. Schnedler (2008): "Too Few Cooks Spoil the Broth: Division of Labour and Directed Production," *The B.E. Journal of Economic Analysis & Policy*, 8(1, Topics), Article 27.

Rogerson, W. P. (1985): "The First-Order Approach to Principal-Agent Problems," *Econometrica*, 53(6), 1357–1367.

Ross, S. A. (1973): "The Economic Theory of Agency: The Principal's Problem," *American Economic Review*, 63(2), 134–39.

Schnedler, W. (2008): "When is It Foolish to Reward for A While Benefiting from B?," *Journal of Labor Economics*, 26(4), 595–619.

——— (2010): "Moral-Hazard, Organization, and Identification," Discussion Paper XX, University of Heidelberg.

Shavell, S. (1979): "Risk-Sharing and Incentives in the Principal-Agent Relationship," *Bell Journal of Economics*, 10(1), 55–73.

Spence, M., and R. Zeckhauser (1971): "Insurance, Information, and Individual Action," *American Economic Review*, 61(2), 380–87.

Strausz, R. (1999): "Efficiency in Sequential Partnerships," *Journal of Economic Theory*, 85(1), 140 – 156.

Tinbergen, J. (1952): *On the Theory of Economic Policy.* North-Holland Publishing Company, Amsterdam.

A Proofs

Proof Proposition 1. Fix a partition $\mathcal{P}$ and a signal structure $\mathcal{S}$. Then, condition (3) is necessary by Lemma 1 in Appendix B. In order to see that it is also sufficient, we show first that for an appropriate choice of linear payments $w^l : \mathbb{R}^k \to \mathbb{R}$, where $w^l(s) = w^l_0 + w^l_1 s_1 + \ldots w^l_k s_k$ for all l, the first-order conditions are met and then that agents' expected utility is concave. First, choose the base payment w^l_0 as an implicit function of $(w^l_1, \ldots, w^l_k)$ such that the outside option is met for agent l at $\mathring{a}$:

$$\mathrm{E}_\epsilon \left[u^l(w^l_0 + w^l_1 S_1 + \ldots + w^l_k S_k, c^l(a)) \right] \Big|_{a=\mathring{a}} = \underline{u}^l. \tag{7}$$

Recall that agent l's first-order conditions are:

$$\begin{aligned} & D_\mu \mathrm{E}_\epsilon \left[u^l(w^l(S(\mu, \epsilon)), c^l) \right] \cdot D_{a^l}\mu(a) \Big|_{a=\mathring{a}} \\ = \; & -\frac{d}{dc^l}\mathrm{E}_\epsilon \left[u^l(w^l(S(\mu, \epsilon)), c^l) \right] \cdot D_{a^l} c^l(a) \Bigg|_{a=\mathring{a}}, \text{ for all } l. \end{aligned} \tag{2}$$

Since condition (3) holds for $\mathring{a}$, there is a λ^l such that

$$\lambda^l D_{a^l}\mu(a)|_{a=\mathring{a}} = D_{a^l}c^l(a)|_{a=\mathring{a}}. \tag{8}$$

Suppose that the following holds:

$$\begin{aligned} &\frac{d}{d\mu_j}\mathrm{E}_\epsilon\left[u^l(w_0^l + w_1^l S_1(\mu_1(a),\epsilon_1) + \ldots + w_k^l S_k(\mu_k(a),\epsilon_k), c^l)\right]\Big|_{a=\mathring{a}} \\ &= \lambda_j^l \cdot \mathrm{E}_\epsilon\left[-\frac{\partial}{\partial c^l}u^l(w_0^l + w_1^l S_1(\mu_1(a),\epsilon_1) + \ldots + w_k^l S_k(\mu_k(a),\epsilon_k), c^l(a))\right]\Big|_{a=\mathring{a}}. \end{aligned} \tag{9}$$

Then, λ^l can be replaced in equation (8), so that the first-order conditions (2) are met at $a = \mathring{a}$. For this to be true, all that remains to be shown is that $(w_1^l,\ldots,w_k^l)$ can be chosen in dependence of λ_j^l such that for all j equation (9) holds. Given (7), the left-hand side in this equation is continuous in $(w_1^l,\ldots,w_k^l)$. Moreover, its absolute value is larger than $|w_j^l|\cdot\kappa_j^l$, for some $\kappa_j^l > 0$ by Lemma 2 in Appendix B. On the other hand, the right-hand side is continuous in $(w_1^l,\ldots,w_k^l)$ and its absolute value is smaller than $|\lambda_j^l|$ because w_0^l is continuous and $0 > \frac{\partial u}{\partial c^l} \geq -1$. Taken together, this ensures the existence of $(\mathring{w}_1^l,\ldots,\mathring{w}_k^l)$ such that (9) is met and the first-order conditions hold—for an illustration with positive λ_j^l see Figure 2.

In the second and last step, it will be shown that first-order conditions are also sufficient because agent l's expected utility is concave in a^l given linear payments. The reason is the following. By definition $S(\mu(a),e)$ is concave in a for almost all realizations e of ϵ. Hence, $w^l(S(\mu(a),e)) = w_0^l + w_1^l S_1 + \ldots + w_k^l S_k$ is concave in a for almost all e. Moreover, $c^l(a)$ is convex in a and $u^l(w^l,c^l)$ is concave in w^l and c^l by definition. Since $u^l(w^l,c^l)$ decreases in c, $u^l(w^l(S(\mu(a),e)),c^l(a)))$ is concave in a for almost all e. The concavity is maintained when integrating over ϵ, so that $\mathrm{E}_\epsilon\left[u^l(w^l(S(\mu(a),\epsilon),c^l(a)))\right]$ is concave in a and consequently also in any sub-vector a^l. □

Proof Proposition 2. First, we show that rank $(D_{a^l}\mu(a)) < |N^l|$ for some l and a implies an identification problem. Let a^l denote the respective activity by agent l. Due to rank$(D_{a^l}\mu(a)) < |N^l|$, the set $\{\hat{a}^l | D_{a^l}\mu(a)\hat{a}^l = (0,\ldots,0)'\}$ is not empty. Since A is an open set, there is a value $\tilde{a}^l$ with $\tilde{a}^l \neq a^l$ but $\mu(\tilde{a}^l,a) = \mu(a^l,a^{-l})$ and the mechanism designer faces an identification problem.

Second, we prove that an identification problem implies rank$(D_{a^l}\mu(a)) < |N^l|$ for some l and a. By definition, an identification problem is present if

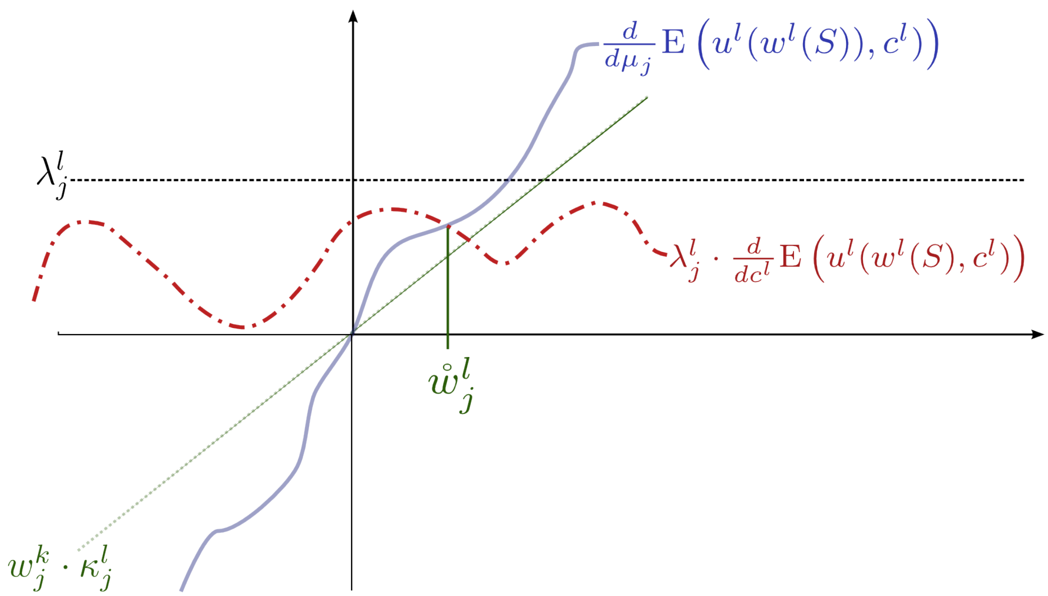

Figure 2: Existence of a payment rate, $\mathring{w}_j^l$, such that agent l's first-order condition are met.

there are differing activities a^l and $\tilde{a}^l$ for some agent l such that $\mu(a^l, a^{-l}) = \mu(\tilde{a}^l, a^{-l})$. Since μ is continuously differentiable and A is convex, there is an activity $\hat{a}^l$ such that $0 = \mu(a^l, a^{-l}) - \mu(\tilde{a}, a^{-l}) = D_{a^l}\mu(\hat{a})(a^l - \tilde{a}^l)$. This, however, implies that $\operatorname{rank}(D_{a^l}\mu(\hat{a})) < |N^l|$. $\square$

Proof Theorem 1. First, we prove that an identification problem leads to limited implementability. Observe that strict convexity of costs $c^l(a)$ and concavity of $\mu(a)$ implies that there is a *unique* least costly way for agent l to produce $\bar{\mu}$:

$$\arg \min_{\hat{a}^l \in \{\check{a}^l | \mu(\check{a}^l, a^{-l}) = \bar{\mu}\}} c^l(\hat{a})|_{\hat{a}=(\hat{a}^l, a^{-l})}.$$

Suppose there is an identification problem, then there is some $\bar{\mu}$ and $\tilde{a}^l \neq a^l$ such that $\bar{\mu} = \mu(a^l, a^{-l}) = \mu(\tilde{a}^l, a^{-l})$. Since the least costly way to produce $\tilde{\mu}$ is unique, either a^l or $\tilde{a}^l$ cannot be induced and implementability is limited.

Second, we show that any activity can be implemented in the absence of identification problems. If there is no identification problem, $\operatorname{rank}(D_{a^l}\mu(a)) \geq |N^l|$ for all l and a by Proposition 2. Take an arbitrary activity a and agent l and let a^l be agent l's decision given a. From $\operatorname{rank}(D_{a^l}\mu) \geq |N^l|$, it follows that $|N^l| \leq k$. Thus, $min(|N^l|, k+1) = |N^l|$ and

$$\text{rank}\begin{pmatrix} D_{a^l}\mu(a) \\ D_{a^l}c^l(a) \end{pmatrix} \leq |N^l|. \qquad (10)$$

On the other hand,

$$\text{rank}\begin{pmatrix} D_{a^l}\mu(a) \\ D_{a^l}c^l(a) \end{pmatrix} \geq \text{rank}(D_{a^l}\mu(a)). \qquad (11)$$

Together, (10) and (11) imply:

$$|N^l| \geq \text{rank}\begin{pmatrix} D_{a^l}\mu(a) \\ D_{a^l}c^l(a) \end{pmatrix} \geq \text{rank}(D_{a^l}\mu(a)) \geq |N^l|,$$

so that

$$\text{rank}\begin{pmatrix} D_{a^l}\mu(a) \\ D_{a^l}c^l(a) \end{pmatrix} = \text{rank}(D_{a^l}\mu(a)).$$

Applying Corollary 2 then yields that a^l can be implemented with agent l. □

Proof Proposition 3. By definition, identification problems due to multi-tasking can be solved by the finest partition of tasks $\{\{1\}, \ldots, \{n\}\}$. Thus, there exists an organization that solves the identification problem. Conversely, if some organization solves the identification problem, there is a partition $\mathcal{P}$ with enough independent signals for each agent l: $\text{rank}(D_{a^l}\mu(a)) \geq |N^l|$ by Proposition 2. Consider partitioning the $|N^l|$ tasks and assigning them to $|N^l|$ agents. Then, for each agent $\tilde{l} \in N^l$, $D_{a^{\tilde{l}}}\mu(a)$ is a $(k,1)-$vector and $\text{rank}(D_{a^l}\mu(a)) \geq |N^l|$ implies $\text{rank}(D_{a^{\tilde{l}}}\mu(a)) = 1$. Since each of the agents only carries out one task $|N^{\tilde{l}}| = 1$ and $\text{rank}(D_{a^{\tilde{l}}}\mu(a)) = 1 = |N^{\tilde{l}}|$, Proposition 2 implies that there is no identification problem. The identification problem thus also disappears with the finest partition.

□

Proof Proposition 4. Suppose there is an identification problem given partition $\mathcal{P}$. Eliminate multi-tasking by considering the finest partition $\{\{1\}, \ldots, \{n\}\}$. The finest partition solves the identification problem if and only if for all l and $\tilde{a}^l \neq a^l$ it follows that $\mu(a^l, a^{-l}) \neq \mu(\tilde{a}^l, a^{-l})$. Since each agent l carries out exactly one task i, this is true if and only if $\mu(\ldots, a_i, \ldots) \neq \mu(\ldots, \tilde{a}_i, \ldots)$ for all i and $a_i \neq \tilde{a}_i$. □

Proof Proposition 5. The proof works by contradiction. Let $\tilde{a} = \arg\min_a \operatorname{rank}(D_a\mu(a))$. Suppose $\tilde{k}m < n$, where $\tilde{k} := \min_a \operatorname{rank}(D_a\mu(a))$. Then,

$$\underbrace{\text{rank } D_a\mu(a)|_{a=\tilde{a}} + \ldots + \text{ rank} D_a\mu(a)|_{a=\tilde{a}}}_{m \text{ summands}} < n = \sum_{l=1}^{m} |N^l|.$$

This implies:

$$\sum_{l=1}^{m} \text{rank } D_{a^l}\mu(a)|_{a=\tilde{a}} < \sum_{l=1}^{m} |N^l|. \tag{12}$$

Now, take an arbitrary partition $\mathcal{P}$ and suppose that there is no identification problem. By Proposition 2, it must then hold that rank $(D_{a^l}\mu(a)) \geq |N^l|$ for all l. This, however, contradicts equation (12). Consequently, the identification problem persists for any partition. □

Proof Proposition 6. Since $\tilde{k} := \min_a \operatorname{rank}(D_a\mu)$, there are at least $\tilde{k}$ independent columns of $D_a\mu$ for all a. Assign the tasks belonging to these columns to the first agent. For this agent, it now holds that rank $(D_{a^1}\mu(a)) \geq \tilde{k} = |N^1|$ for all a and his activities are identified. Since $m > n - \tilde{k}$, there remain at least $n - \tilde{k}$ agents who are not yet assigned to tasks. Assign each of these agents to one of the remaining $n - \tilde{k}$ tasks. Since the identification problem is caused by multi-tasking, the activities of these agents are also identified. □

B Auxiliary Results

Lemma 1 (Necessary condition for implementation)**.** *Given signal structure $\mathcal{S}$ and partition $\mathcal{P}$, an activity $\mathring{a}$ can only be implementable if the following holds:*

$$\lambda^l \, D_{a^l}\mu(a)|_{a=\mathring{a}} = D_{a^l}c^l(a)\Big|_{a=\mathring{a}} \text{ with some } \lambda^l \in \mathbb{R}^k \text{ for all } l. \tag{3}$$

Proof. Since we are looking for an inner maximizer, the first-order conditions are necessary. Using the separability assumption, these conditions are equivalent to (2). Next suppose that Condition (3) is violated and show that first-order conditions cannot be met then. Suppose for all l and $\lambda^l : \lambda^l D_{a^l}\mu(a)|_{a=\mathring{a}} \neq D_{a^l}c^l(a)|_{a=\mathring{a}}$. Now choose $\tilde{\lambda}^l := -\frac{D_\mu \mathrm{E}_\epsilon\left[u^l(w^l(S(\mu,\epsilon)),c^l)\right]}{\frac{d}{dc^l}\mathrm{E}_\epsilon\left[u^l(w^l(S(\mu,\epsilon)),c^l)\right]}\Bigg|_{a=\mathring{a}}$, where the denominator is strictly negative because of $\frac{\partial u^l}{\partial c^l} < 0$. Then,

$$\tilde{\lambda}^l D_{a^l}\mu(a) \neq D_{a^l}c^l(a),$$

which implies directly that the first-order condition must be violated, too. $\square$

Lemma 2. *Fix μ, and c, and and consider linear payments $w^l(\cdot)$ such that the agents' outside options are met. Then, there exists a $\kappa_j^l > 0$ such that:*

$$\left|\frac{d}{d\mu_j}E_\epsilon\left[u^l(w^l(S),c)\right]\right|_{a=\mathring{a}} \geq |w_j^l|\cdot\kappa_j^l. \tag{13}$$

Proof. Define $S_{-j} := (S_1,\ldots,S_{j-1},S_{j+1},\ldots,S_k)$ and let F_{-j} be the respective cumulative distribution. Denote the c.d.f. of S_j given S_{-j} by $F_j(s_j|\mu_j)$ and the respective density by $f_j(s_j|\mu)$. Let us focus on a continuously distributed $S_j(\mu_j,\epsilon_j)$—the analysis in the case of a discrete distribution is analogous. For given costs c, the gain in expected utility from a change from μ_j to some $\tilde{\mu}_j$ with $\tilde{\mu}_j > \mu_j$ can be written as:

$$\begin{aligned}
&\mathrm{E}_S\left[u^l(w^l(S),c)|(\mu_1,\ldots,\mu_{j-1},\tilde{\mu}_j,\mu_{j+1},\ldots,\mu_k)\right] - \mathrm{E}_S\left[u^l(w^l(S),c)|\mu\right] \\
&= \iint u^l(w_0^l+\ldots+w_j^l s_j+\ldots+w_k^l s_k,c)\cdot\left[f_j(s_j|\tilde{\mu}_j)-f_j(s_j|\mu_j)\right]ds_j\,dF_{-j} \\
&= \left[u^l(w^l(s_j),c)(F_j(s_j|\tilde{\mu}_j)-F_j(s_j|\mu_j))\right]_{\underline{s}_j}^{\overline{s}_j} \qquad (14) \\
&\quad -\iint \frac{\partial}{\partial w}u^l(w^l(s),c)\cdot w_j\cdot\left[F_j(s_j|\tilde{\mu}_j)-F_j(s_j|\mu_j)\right]ds_j\,dF_{-j},
\end{aligned}$$

where $\underline{s}_j$ and $\overline{s}_j$ are the bounds of the support of S_j and possibly $-\infty$ and $+\infty$. The last equality follows from integration by parts. Moreover the term in (14) is zero because both c.d.f.'s are zero evaluated at the lower bound and one at the upper bound. Dividing both sides by $\tilde{\mu}_j - \mu_j$ and letting $\tilde{\mu}_j \to \mu_j$, we obtain the derivative of the agent's expected utility with respect to μ_j for a given c:

$$\begin{aligned}
&\frac{d}{d\mu_j}\mathrm{E}_S\left[u^l(w_0^l + w_1^l S_1 + \ldots + w_j^l S_j + \ldots + w_k^l S_k, c)\right] \\
&= -\iint \frac{\partial}{\partial w}u^l(w^l(s), c) \cdot w_j^l \cdot \frac{\partial}{\partial \mu_j}F_j(s_j|\mu_j)\; ds_j\, dF_{-j} \\
&= -\iint \frac{\partial}{\partial w}u^l(w^l(s), c) \cdot w_j \cdot f_j(s_j|\mu_j) \cdot \frac{\frac{\partial}{\partial \mu_j}F_j(s_j|\mu_j)}{f_j(s_j|\mu_j)}\; ds_j\, dF_{-j} \\
&= w_j^l \cdot \mathrm{E}_S\left[\frac{\partial}{\partial w}u^l(w^l(S), c) \cdot \frac{-\frac{\partial}{\partial \mu_j}F_j(S_j|\mu_j)}{f_j(S_j|\mu_j)}\right] \\
&= w_j^l \cdot \mathrm{E}_S\left[\frac{\partial}{\partial w}u^l(w(S), c)\right] \cdot \gamma_j^l \text{ with } \gamma_j^l > 0,
\end{aligned} \tag{15}$$

where the last equality follows from Lemma 3. Given that the agent l's outside option is met, Lemma 4 implies:

$$\mathrm{E}_S\left[\frac{\partial}{\partial w}u^l(w(S), c)\right] \geq \delta^l > 0$$

Using this in (15) yields:

$$\left|\frac{d}{d\mu_j}\mathrm{E}_\epsilon\left[u^l(w^l(S), c)\right]\right|_{a=\mathring{a}} \geq |w_j^l| \cdot \kappa_j^l \text{ with } k_j^l := \delta^l \cdot \gamma_j^l > 0.$$

□

Lemma 3. *Given μ, c, there is a strictly positive γ_j for all payments $w(\cdot)$ such that:*

$$E_S\left[\frac{\partial}{\partial w}u^l(w(S), c) \cdot \frac{-\frac{\partial}{\partial \mu_j}F_j(S_j|\mu_j)}{f_j(S_j|\mu_j)}\right] = E_S\left[\frac{\partial}{\partial w}u^l(w(S), c)\right] \cdot \gamma_j^l,$$

where $F_j(s_j|\mu_j) = Prob(S_j \leq s_j | S_{j'} = s_{j'}$ for all $j' \neq j)$ and $f_j(s_j) = \frac{\partial}{\partial s_j}F_j(s_j)$.

Proof. Define the random variable $B_j := \frac{-\frac{\partial}{\partial \mu_j}F_j(S_j|\mu_j)}{f_j(S_j|\mu_j)}$. Using this notation, we get:

$$
\begin{aligned}
& \mathrm{E}_S\left[\frac{\partial}{\partial w}u^l(w(S),c)\cdot\frac{-\frac{\partial}{\partial \mu_j}F_j(S_j|\mu_j)}{f_j(S_j|\mu_j)}\right] \\
= & \quad \mathrm{E}\left[\frac{\partial}{\partial w}u^l(w(S),c)\cdot B_j\,\middle|B_j\geq\beta_j\right]\cdot Prob(B_j\geq\beta_j) \\
+ & \quad \mathrm{E}_S\left[\frac{\partial}{\partial w}u^l(w(S),c)\cdot B_j\,\middle|B_j<\beta_j\right](1-Prob(B_j\geq\beta_j)) \\
\geq & \quad \mathrm{E}_S\left[\frac{\partial}{\partial w}u^l(w(S),c)\cdot \beta_j\,\middle|B_j\geq\beta_j\right]\cdot Prob(B_j\geq\beta_j) \\
+ & \quad \mathrm{E}_S\left[\frac{\partial}{\partial w}u^l(w(S),c)\cdot 0\quad\middle|B_j<\beta_j\right](1-Prob(B_j\geq\beta_j)) \\
= & \quad \mathrm{E}_S\left[\frac{\partial}{\partial w}u^l(w(S),c)\right]\cdot\gamma_j, \qquad (16)
\end{aligned}
$$

where the existence of $\gamma_j \in [0,\beta_j]$ follows from the intermediate value theorem. Moreover, the event $B_j \geq \beta_j$ occurs with positive probability for some $\beta_j > 0$ because S_j is stochastically increasing in μ_j, so that:

$$Prob(B_j>\beta_j) = Prob\left(\frac{-\frac{\partial}{\partial \mu_j}F_j(s_j|\mu_j)}{f_j(s_j|\mu_j)}\geq\beta_j\right)>0.$$

This, however, means that γ_j has to be strictly larger than zero. □

Lemma 4. *Take a μ, c, and payments $w(\cdot)$ and $\tilde{w}(\cdot)$ such that*

$$E\left[u^l(w(S(\mu,\epsilon),c))\right] = E\left[u^l(\tilde{w}(S(\mu,\epsilon),c))\right].$$

Then, for some positive δ^l :

$$E\left[\frac{\partial}{\partial w}u^l(w(S(\mu,\epsilon)),c)\right]\geq\delta^l>0.$$

Proof. By Lemma 5 there exists a

$\hat{w}^{\min} := \min\{\hat{w}|Prob(w(S(\mu,\epsilon))\leq\hat{w}) = Prob(\tilde{w}(S(\mu,\epsilon))\leq\hat{w})>0\}\subseteq[\underline{w},\overline{w}].$

Using this definition, we get:

$$
\begin{aligned}
& \mathrm{E}_\epsilon\left[\frac{\partial}{\partial w}u^l(w(S(\mu,\epsilon)),c)\right] \\
= & \quad \mathrm{E}\left[\frac{\partial}{\partial w}u^l(w(S(\mu,\epsilon)),c)\middle| w(S(\mu,\epsilon))\le \hat{w}^{\min}\right]\cdot Prob(w(S(\mu,\epsilon))\le \hat{w}^{\min}) \\
& + \mathrm{E}\left[\frac{\partial}{\partial w}u^l(w(S(\mu,\epsilon)),c)\middle| w(S(\mu,\epsilon)) > \hat{w}^{\min}\right]\cdot Prob(w(S(\mu,\epsilon)) > \hat{w}^{\min}) \\
\ge & \quad \mathrm{E}\left[\frac{\partial}{\partial w}u^l(w(S(\mu,\epsilon)),c)\middle| w(S(\mu,\epsilon))\le \hat{w}^{\min}\right]\cdot Prob(w(S(\mu,\epsilon))\le \hat{w}^{\min}),
\end{aligned}
$$

where $Prob(w(S(\mu,\epsilon)) \le \hat{w}^{\min}) > 0$ by the definition of $\hat{w}^{\min}$. Since u^l is non-convex, $\frac{\partial}{\partial w}u^l(w,c)$ weakly falls in w so that

$$
\mathrm{E}\left[\frac{\partial}{\partial w}u^l(w(S(\mu,\epsilon)),c)\middle| w(S(\mu,\epsilon))\le \hat{w}^{\min}\right] \ge \left.\frac{\partial}{\partial w}u^l(w,c)\right|_{w=\hat{w}^{\min}},
$$

which is strictly larger than zero because $\frac{\partial}{\partial w}u^l(w,c) > 0$. Taken together, this implies:

$$
\mathrm{E}_\epsilon\left[\frac{\partial}{\partial w}u^l(w(S(\mu,\epsilon)),c)\right] \ge \underbrace{\left.\frac{\partial}{\partial w}u^l(w,c)\right|_{w=\hat{w}^{\min}}\cdot Prob(w(S(\mu,\epsilon))\le \hat{w}^{\min})}_{=:\delta^l} > 0.
$$

□

Lemma 5. *Take a μ, and c and consider payments $w(\cdot)$ and $\tilde{w}(\cdot)$ such that:*

$$
E\left[u(w(S(\mu,\epsilon),c))\right] = E\left[u(\tilde{w}(S(\mu,\epsilon),c))\right].
$$

Then, there are finite numbers $\underline{w} < \overline{w}$ in the support such that for all $w(\cdot)$ and $\tilde{w}(\cdot)$, there is a $\hat{w} \in [\underline{w},\overline{w}]$ in the support of of $w(S(\mu,\epsilon))$ and $\tilde{w}(S(\mu,\epsilon))$ with

$$
Prob(w(S(\mu,\epsilon)) \le \hat{w}) = Prob(\tilde{w}(S(\mu,\epsilon))) \le \hat{w}).
$$

Proof. The proof works by contradiction. If for all $\overline{w}$ there exist $w(\cdot)$ and $\tilde{w}(\cdot)$, such that for all $\hat{w} \le \overline{w}$ it holds that

$$
Prob(w(S(\mu,\epsilon)) \le \hat{w}) < Prob(\tilde{w}(S(\mu,\epsilon)) \le \hat{w}),
$$

then $w(S(\mu,\epsilon))$ first-order stochastically dominates $\tilde{w}(S(\mu,\epsilon))$ for values below $\overline{w}$. Together with the assumption that u^l increases in w, this yields:

$$\mathrm{E}\left[u^l(w(S(\mu,\epsilon),c))|w(S(\mu,\epsilon))\leq\overline{w}\right] > \mathrm{E}\left[u^l(\tilde{w}(S(\mu,\epsilon),c))|\tilde{w}(S(\mu,\epsilon))\leq\overline{w}\right].$$

Since this holds for all $\overline{w}$, it also holds if $\overline{w}$ approaches infinity, which means that $Prob(w(S(\mu,\epsilon))\leq\overline{w}) = Prob(\tilde{w}(S(\mu,\epsilon))\leq\overline{w})$ becomes one, so that

$$\begin{aligned}
\mathrm{E}\left[u^l(w(S(\mu,\epsilon),c))\right] &= \mathrm{E}\left[u^l(w(S(\mu,\epsilon),c))|w(S(\mu,\epsilon))\leq\overline{w}\right] \\
&> \mathrm{E}\left[u(\tilde{w}(S(\mu,\epsilon),c))|\tilde{w}(S(\mu,\epsilon))\leq\overline{w}\right] \\
&= \mathrm{E}\left[u(\tilde{w}(S(\mu,\epsilon),c))\right], \qquad (17)
\end{aligned}$$

which contradicts $\mathrm{E}\left[u(w(S(\mu,\epsilon),c))\right] = \mathrm{E}\left[u(\tilde{w}(S(\mu,\epsilon),c))\right]$. Analogously, one can show that $\hat{w}$ is above some finite lower bound $\underline{w}$. □

C Additional Results

Lemma 6 (Dual representation)**.** *Let $Y \sim G_{\mu_j}(y)$, where G is the cumulative distribution function of Y and μ_j a real-valued parameter (Mirrlees representation).*[16] *Alternatively, consider the state-space representation from Section 1, which involves S_j, μ_j and ϵ_j. Then, there is a state-space representation if and only if there is a Mirrlees representation.*

Proof. First, show that a state-space representation for $Y \sim G_{\mu_j}(y)$ exists. Define $G^{-1}_{\mu_j}: \mathbb{R}\to\mathbb{R}$ with:

$$G^{-1}_{\mu_j}(e_j) = \min\left\{y|G_{\mu_j}(y) = \max\left\{\tilde{p}|\tilde{p} = G_{\mu_j}(\tilde{y})\leq e_j \text{ for some } \tilde{y}\right\}\right\}.$$

Now, set $S_j(\mu_j,\epsilon_j) := G^{-1}_{\mu_j}(\epsilon_j)$. Next, take $\epsilon_j \sim \mathrm{UNIF}[0;1]$. Then,

$$\mathrm{Prob}(S_j(\mu_j,\epsilon_j)\leq y) = \mathrm{Prob}(G^{-1}_{\mu_j}(\epsilon_j)\leq y) = G_{\mu_j}(y).$$

To find a Mirrlees representation, set $Y := S_j(\mu_j,\epsilon_j)$ and $G_{\mu_j}(y) = \mathrm{Prob}(Y \leq y) = \mathrm{Prob}(S_j(\mu_j,\epsilon_j)\leq s)$. □

Proposition 7 (Identification and implementability with limited liability)**.** *Consider a risk-neutral agent with utility, $u(w^l,c^l) = w^l - c^l$, and limited liability. Then, any generic activity $\mathring{a} = (a^l, a^{-l})$ can be implemented with a finite bonus payment, if there is no identification problem.*

[16] The terminology is based on Conlon (2009).

Proof. There is no identification problem, so $\mu(\mathring{a}^l, \mathring{a}^{-l}) \neq \mu(\tilde{a}^l, \mathring{a}^{-l})$ for all $\mathring{a}^l \neq \tilde{a}^l$ and l. Consequently, the distribution of $S(\mu(\mathring{a}^l, \mathring{a}^{-l}), \epsilon)$ differs from that of $S(\mu(\tilde{a}^l, \mathring{a}^{-l}), \epsilon)$. In particular, there must be some set S^1 such that for all $\tilde{a}^l$: $Prob\left(S(\mu(\mathring{a}^l, \mathring{a}^{-l}), \epsilon) \in S^1\right) > Prob\left(S(\mu(\tilde{a}^l, \mathring{a}^{-l}), \epsilon) \in S^1\right)$. Define the first probability as $P(\mathring{a}^l)$ and the latter as $P(\tilde{a}^l)$. Now, pay a bonus $w^l \geq 0$ whenever $s \in S^1$ and nothing otherwise. Then, the incentive constraint of agent l becomes $w^l P(\mathring{a}^l) - c^l(\mathring{a}^l, \mathring{a}^{-l}) \geq w^l P(\tilde{a}^l) - c^l(\tilde{a}^l, \mathring{a}^{-l})$. This inequality holds if the bonus w^l is large enough: $w^l \geq \frac{c^l(\mathring{a}^l, \mathring{a}^{-l}) - c^l(\tilde{a}^l, \mathring{a}^{-l})}{P(\mathring{a}^l) - P(\tilde{a}^l)}$. The lower bound is finite for $\mathring{a}^l \neq \tilde{a}^l$. Moreover, for $\tilde{a}^l \to \mathring{a}^l$, it converges to the finite real number $D_a c^l(a)\hat{a}/D_a P(a)\hat{a}$, where $\hat{a}$ is the direction from which $\tilde{a} = (\tilde{a}^l, a^{-l})$ approaches $\mathring{a}$. Accordingly, the desired choice $\mathring{a}^l$ can be implemented with a finite bonus payment, w^l. $\square$