Supplement to "A unifying approach to incentive compatibility in moral hazard problems"

(*Theoretical Economics*, Vol. 12, No. 1, January 2017, 25–51)

René Kirkegaard
Department of Economics and Finance, University of Guelph

Appendix A: Properties of the likelihood ratio

A.1 *Interpretation of* $l_a(\mathbf{x}|a)$ Δ*-antitone*

The restriction that $l_a(\mathbf{x}|a)$ is Δ-antitone or Δ^2-antitone have implications for how the dependence between signals must change with a. In particular, for the likelihood ratio to be Δ-antitone it is necessary that

$$\frac{\partial^2}{\partial x_i \partial x_j}\left(\frac{\partial \ln f(\mathbf{x}|a)}{\partial a}\right) = \frac{\partial}{\partial a}\left(\frac{\partial^2 \ln f(\mathbf{x}|a)}{\partial x_i \partial x_j}\right) \leq 0 \quad \text{for all } i, j, i \neq j. \tag{A.1}$$

Recall that the random variables are said to be (positively) affiliated if the term inside the last parentheses is always nonnegative, while they are "negatively" affiliated if the term is always nonpositive. See Milgrom and Weber (1982) and Müller and Stoyan (2002). Thus, the sign determines whether variables are positively or negatively dependent. Technically, the numerical magnitude of the term inside the parentheses does not appear to have any formal interpretation, but it is nevertheless noteworthy that (A.1) requires it to be decreasing in a. This suggests that the variables become "less dependent" (or "less affiliated") as a increases. In particular, if the variables are independent for some $a = \widehat{a}$, or $\partial^2 \ln f(\mathbf{x}|\widehat{a})/\partial x_i \partial x_j = 0$, then the variables must be positively affiliated for $a < \widehat{a}$ and negatively affiliated for $a > \widehat{a}$, by (A.1).

The weakening dependence may suggest that it becomes less likely that all signals are large simultaneously. However, recall that MLRP together with affiliation is sufficient to ensure NISP. Then, as long as the signals remain affiliated, an increase in a causes the signals to exhibit less dependence while nevertheless making it more likely that they are all large (i.e., in any given increasing set) simultaneously.

Consistent with Corollaries 1 and 2, (A.1) suggest that the restriction that $l_a(\mathbf{x}|a)$ is Δ-antitone has little bite if the dependence structure is the same for all a. For instance, consider n jointly normally distributed signals, where a determines the mean values, $\mu_i(a)$, with $\mu_i'(a) > 0$. As long as the covariance matrix is independent of a, $l_a(\mathbf{x}|a)$ is then Δ-antitone and Δ^2-antitone if and only if MLRP is satisfied. However, since $l_a(\mathbf{x}|a)$ is not

René Kirkegaard: rkirkega@uoguelph.ca

DOI: 10.3982/TE2008

bounded below, this distribution violates one of the other assumptions of the model. Nevertheless, if the likelihood ratio of a parent distribution is Δ-antitone or Δ^2-antitone, then the likelihood ratio maintains those properties upon truncation.[1] In other words, the FOA can be justified as long as the truncated normal distribution satisfies only MLRP and CLOCC.

For another example, assume that the joint density takes the form

$$f(\mathbf{x}|a,\theta) = e^{\sum_{i=1}^{n}[h^i(x_i,a)+r^i(a)]+\theta t(\mathbf{x})+r(a,\theta)} = \left(\prod_{i=1}^{n} e^{h^i(x_i,a)+r^i(a)}\right) \times e^{\theta t(\mathbf{x})+r(a,\theta)},$$

where θ is a parameter that measures the dependence between the signals.[2] Note that the signals are independent if $\theta = 0$. Taking h^i as given, r^i is determined such that the ith term in the parentheses integrates to 1. Taking $t(\mathbf{x})$ as given, the term $r(a,\theta)$ is determined such that $f(\mathbf{x}|a,\theta)$ integrates to 1. Thus, $r(a,0) = 0$ for all a.

The likelihood ratio is Δ-antitone if $h^i_{ax} \geq 0$ and Δ^2-antitone if it is also the case that $h^i_{axx} \leq 0$. These are the same conditions that ensure that $f^i(x_i|a,0)$ has a monotonic and a concave likelihood ratio, respectively. Moreover, θ is irrelevant for these properties. Thus, in this model interdependence does not make it any harder to satisfy the conditions on the likelihood ratio in Propositions 3 or 4. Furthermore, it is possible to allow θ to be a function of a if more structure is imposed on $t(\mathbf{x})$. For example, if $t(\mathbf{x})$ is supermodular, then $l_a(\mathbf{x}|a)$ is Δ-antitone in the bivariate case if MLRP is satisfied and $\theta'(a) \leq 0$. Note that in this case signals are positively (negatively) affiliated if $\theta(a)$ is positive (negative).

A.2 *Monotonically decreasing Fisher information*

It is clear from Section 6 that there is an important relationship between Fisher information and the second-best action in the SQIT model. It is relevant whether $V(a)$ is increasing, decreasing, or possibly non-monotonic in a. However, at the intuitive level it may be hard to get a feel for how Fisher information depends on a. Indeed, the distribution function used in Example 1 in Jewitt et al. (2008) can be shown to exhibit U-shaped Fisher information on $[\underline{a},\overline{a}]$.[3] Alternatively, as mentioned earlier, Holmström (1979) presents a one-signal example in which x is exponentially distributed, $F(x|a) = 1 - e^{-x/a}$, $a > 0$, $x \geq 0$. The Fisher information in this case is a^{-2}, which is decreasing in a.[4]

[1] Let $G(\mathbf{x}|a)$ denote the parent distribution, with density function $g(\mathbf{x}|a)$. Since $f(\mathbf{x}|a) = g(\mathbf{x}|a)/\int_{\underline{\mathbf{x}}}^{\overline{\mathbf{x}}} g(\mathbf{y}|a)\,d\mathbf{y}$ it holds that $\ln f(\mathbf{x}|a) = \ln g(\mathbf{x}|a) - \ln \int_{\underline{\mathbf{x}}}^{\overline{\mathbf{x}}} g(\mathbf{y}|a)\,d\mathbf{y}$. The claim now follows from the fact that the last term is independent of $\mathbf{x}$.

[2] This distribution is related to a distribution used in Banker and Datar's (1989) paper on linear aggregation of signals.

[3] Their example is $f(x|a) = 1 + \frac{1}{2}(1-2x)(1-2a)$, $x, a \in [0,1]$. Similarly, the Fisher information of a Bernoulli distribution with $a \in (0,1)$ being the probability of a success (and $1-a$ the probability of a failure) is $V(a) = 1/a(1-a)$, which is also U-shaped.

[4] Likewise, the Fisher information of a normally distributed variable with mean $\mu(a)$ is $\mu'(a)^2/\sigma^2$. This is decreasing in a as long as $\mu(a)$ is increasing and concave.

Sufficient conditions for Fisher information to be decreasing in a are derived next. To do so, I ask what added structure must be imposed on top of the structure already implied by Rogerson's (1985) or Jewitt's (1988) conditions in the one-signal case. In Rogerson's case, it is sufficient to add the condition that $l_{aax} \leq 0$. In Jewitt's case, $l_{aax} \leq 0 \leq l_{aaxx}$ is sufficient. Note that the likelihood ratio, $l_a(x|a)$, is submodular in x and a when $l_{aax} \leq 0$. Holmström's example satisfies both conditions.[5]

LEMMA A.1 (Monotonic Fisher information). *We have $V'(a) \leq 0$ under either of the following sets of conditions:*

(i) Assume $F(x|a)$ satisfies CDFC and MLRP. Then $V'(a) \leq 0$ if $l_{aax}(x|a) \leq 0$ for all x and a.

(ii) Assume $\int_{\underline{x}}^{x} F_{aa}(y|a)\,dy \geq 0$ for all x and a and that $l_a(x|a)$ is increasing and concave in x for all a. Then $V'(a) \leq 0$ if $l_{aax}(x|a) \leq 0 \leq l_{aaxx}$ for all x and a.

The lemma follows from integration by parts.

Lemma A.1 plays no direct role in the paper. The intention is simply to suggest that results relying on $V'(a) \leq 0$ are perhaps of more interest than results that are based on $V'(a) > 0$.

APPENDIX B: THE SQIT MODEL

This section characterizes the optimal contract in the SQIT model. Thus, consider a pure multitasking environment, i.e., $F(\mathbf{x}|\mathbf{a}) = G^1(\mathbf{x}^1|a_1)G^2(\mathbf{x}^2|a_2)\cdots G^m(\mathbf{x}^m|a_m)$, and assume moreover that $v(w) = 2\sqrt{w}$ or $\omega(z) = 2z$ (Example 1). Assuming the FOA is valid and the contract takes the form in (1), the agent's expected utility can be derived by integrating (2) over all $\mathbf{x}$, yielding

$$EU(a^*) = \int \left(2\lambda + 2\sum_{j=1}^{m} \mu_j l^j_{a_j}(\mathbf{x}^j|a^*_j)\right) f(\mathbf{x}|\mathbf{a}^*)\,d\mathbf{x} - c(\mathbf{a}^*).$$

Since the expected value of $l^j_{a_j}(\mathbf{x}^j|a^*_j)$ is zero, the participation constraint yields $\lambda = \frac{1}{2}(\overline{u} + c(\mathbf{a}^*))$. Similarly, L-IC requires

$$\int \left(2\lambda + 2\mu_j l^j_{a_j}(\mathbf{x}^j|a^*_j) + 2\sum_{k\neq j} \mu_k l^k_{a_k}(\mathbf{x}^k|a^*_k)\right) g^j_{a_j}(\mathbf{x}^j|a^*_j) \prod_{k\neq j} g^k(\mathbf{x}^k|a^*_k)\,d\mathbf{x} - c_j(\mathbf{a}^*) = 0$$

for all $j = 1, 2, \ldots, m$. Since $g^j_{a_j}(\mathbf{x}^j|a^*_j)$ integrates to zero and the expected value of $l^k_{a_k}(\mathbf{x}^k|a^*_k)$ is zero, L-IC simplifies to

$$2\mu_j V_j(a^*_j) - c_j(\mathbf{a}^*) = 0, \quad j = 1, 2, \ldots, m,$$

[5] However, the support of the exponential distribution is unbounded and so Lemma A.1 technically does not apply. LiCalzi and Spaeter (2003) characterize classes of distribution functions that satisfy CDFC and MLRP. Among their explicit examples are $F(x|a) = xe^{a(x-1)}$, $x \in [0,1]$, and $G(x|a) = x + (x - x^2)/(a+1)$, $x \in [0,1]$, with $a > 0$. Both satisfy $l_{aax} \leq 0$.

where

$$V_j(a_j^*) = \int (l^j_{a_j}(\mathbf{x}^j|a_j^*))^2 g^j(\mathbf{x}^j|a_j^*)\, d\mathbf{x}$$

is the Fisher information, measuring how informative $\mathbf{x}^j$ is about a_j. Thus, L-IC yields $\mu_j = \frac{1}{2}c_j(\mathbf{a}^*)/V_j(a_j^*) > 0$.

The agent's utility from income when the signal is $\mathbf{x}$ is now

$$v(w(\mathbf{x})) = \overline{u} + c(\mathbf{a}^*) + \sum_{j=1}^{m}\left(\frac{c_j(\mathbf{a}^*)}{V_j(a_j^*)} l^j_{a_j}(\mathbf{x}^j|a_j^*)\right). \tag{B.1}$$

Thus, $v(w(\mathbf{x}))$ inherits its curvature properties from the likelihood ratios. The expected cost, $K(\mathbf{a}^*)$, to the principal of inducing $\mathbf{a}^*$ can be computed. To begin, note that

$$w(\mathbf{x}) = \left(\frac{\overline{u} + c(\mathbf{a}^*)}{2} + \sum_{j=1}^{m}\left(\frac{c_j(\mathbf{a}^*)}{2V_j(a_j^*)} l^j_{a_j}(\mathbf{x}^j|a_j^*)\right)\right)^2$$

or

$$w(\mathbf{x}) = \left(\frac{\overline{u} + c(\mathbf{a}^*)}{2}\right)^2 + \left(\sum_{j=1}^{m}\left(\frac{c_j(\mathbf{a}^*)}{2V_j(a_j^*)} l^j_{a_j}(\mathbf{x}^j|a_j^*)\right)\right)^2 + 2\left(\frac{\overline{u} + c(\mathbf{a}^*)}{2}\right)\sum_{j=1}^{m}\left(\frac{c_j(\mathbf{a}^*)}{2V_j(a_j^*)} l^j_{a_j}(\mathbf{x}^j|a_j^*)\right).$$

It is necessary to integrate $w(\mathbf{x})$ to derive $K(\mathbf{a}^*)$. The first term in $w(\mathbf{x})$ is a constant, while the third term integrates to zero as the expected value of $l^j_{a_j}(\mathbf{x}^j|a_j^*)$ is zero. Expanding the second term yields terms that involve $(l^j_{a_j}(\mathbf{x}^j|a_j^*))^2$ and $l^j_{a_j}(\mathbf{x}^j|a_j^*)l^k_{a_k}(\mathbf{x}^k|a_k^*)$, $k \neq j$. The former integrates to $V_j(a_j^*)$, while the latter integrates to zero. Thus,

$$K(\mathbf{a}^*) = \int w(\mathbf{x}) f(\mathbf{x}|\mathbf{a}^*)\, d\mathbf{x} = \left(\frac{\overline{u} + c(\mathbf{a}^*)}{2}\right)^2 + \sum_{j=1}^{m}\frac{c_j(\mathbf{a}^*)^2}{4V_j(a_j^*)}. \tag{B.2}$$

The first term coincides with the full-information cost of dictating $\mathbf{a}^*$. The remaining terms thus capture the agency costs under asymmetric information.

Appendix C: Pure multitasking

The standard moral hazard model assumes that the agent's action is one-dimensional. This simplification rules out many interesting and realistic principal–agent relationships, such as those where the agent is assigned multiple tasks. However, the approach suggested here is conceptually robust to such an extension: the isomorphism described in Section 3 remains valid when $\mathbf{a}$ is multidimensional. In fact, it is straightforward to generalize CISP, LOCC, and CLOCC to higher dimensions.

In the one-dimensional case, CISP, LOCC, and CLOCC all exploit the fact that the tangent line to a convex (concave) function is always below (above) the function itself. This property extends to higher dimensions, where the tangent line is replaced by the tangent plane. Thus, for multidimensional $\mathbf{a}$, CISP requires that $P(\mathbf{x} \in \mathbf{E}|\mathbf{a})$ is concave in $\mathbf{a}$ for all increasing sets. Likewise, LOCC becomes the requirement that $F(\mathbf{x}|\mathbf{a})$ is convex in $\mathbf{a}$ for all $\mathbf{x}$. CLOCC is extended in a similar manner: $\int_{\mathbf{y}\leq\mathbf{x}} F(\mathbf{y}|\mathbf{a})\, d\mathbf{y} \geq 0$ must be convex in $\mathbf{a}$ for all $\mathbf{x}$. Then CISP, LOCC, and CLOCC imply, as before, that F^L FOSD F or that F^L dominates F in the lower orthant order or the lower orthant-concave order, respectively.

Unfortunately, it is harder to establish that $v(w(\mathbf{x}))$ has the properties necessary to invoke integral stochastic orders. For example, MLRP does not seem sufficient to even ensure that $v(w(\mathbf{x}))$ is monotonic in the general case. The underlying reason is that it is hard to sign the multipliers when several action dimensions simultaneously impact the distribution of a given signal.[6,7]

To overcome this technical difficulty, I focus on pure multitasking environments, defined in Section 7. The main purpose is to show that in such environments, the FOA is valid under conditions that are analogous to those in Section 4.

The independence assumption embodied in the definition of pure multitasking environments may at first sight appear fairly strong. However, consider for example a professor who invests effort in research and teaching. The professor's research efforts mean that he understands the literature well and has first hand experience with some of the technical problems that can arise in his field. Armed with this knowledge and experience, it becomes easier (e.g., less costly in terms of preparation time) to teach up to a given standard. Whether the professor was "lucky" enough to prove his theorem and earn a top publication is arguably not what drives teaching quality; rather, it is the very effort itself (and not its outcome) that has a spill-over effect. In this case, the tasks are interdependent only through the cost function.

To see why the pure multitasking environment is particularly tractable, recall from Section 7 that the likelihood ratios are separable, or $l_{a_j}(\mathbf{x}|\mathbf{a}) = l^j_{a_j}(\mathbf{x}^j|a_j)$. With the aim of eventually invoking Lemma 2 in Section 4, it is useful to note that $l_{a_j}(\mathbf{x}|\mathbf{a})$ is Δ-antitone in $\mathbf{x}$ if and only if $l^j_{a_j}(\mathbf{x}^j|a_j)$ is Δ-antitone in $\mathbf{x}^j$. Thus, it is sufficient to check the properties of the m task-specific marginal distributions, $G^j(\mathbf{x}^j|a_j)$. Likewise, $v(w(\mathbf{x}))$ simplifies to

$$v(w(\mathbf{x})) = \omega\left(\lambda + \sum_{j=1}^{m} \mu_j l^j_{a_j}(\mathbf{x}^j|a_j^*)\right). \tag{C.1}$$

The separability in (C.1) is key to signing the multipliers. However, it does not seem straightforward to generalize Jewitt's or Rogerson's proof that $\mu > 0$ to allow $m > 1$. Thus,

[6] To illustrate the ensuing complications, assume that there are two action dimensions and that both impact the distribution of x_1. Assume $l_{a_j}(\mathbf{x}|\mathbf{a})$ is strictly increasing in x_1, $j = 1, 2$. Imagine that $\mu_1 > 0 > \mu_2$. Then (2) permits the possibility that $v(w(\mathbf{x}))$ is increasing in x_1 most of the time, but not all the time. Thus, L-IC may very well be satisfied, but G-IC cannot be established using the current approach (or the standard approaches) since the contract is not globally increasing.

[7] Boyer and Sinclair-Desgagné (2001) list a set of assumptions that validates the FOA in multi-task settings. However, they *assume* that the endogenous multipliers are strictly positive. In addition to MLRP they also impose assumptions that are stronger than NISP and CISP. See also Sinclair-Desgagné (1999).

I use a more primitive technique, inspired by an observation in Rogerson (1985, footnote 8) for the $n = m = 1$ case. Recall that $G^j(\mathbf{x}^j|a_j)$ satisfies MLRP if $l^j_{a_j}(\mathbf{x}^j|a_j)$ is non-decreasing in every element of $\mathbf{x}^j$. It satisfies NISP if $P_{a_j}(\mathbf{x}^j \in \mathbf{E}|a_j) \geq 0$ for all increasing sets. MLRP implies NISP if $\mathbf{x}^j$ is one-dimensional, but not necessarily otherwise.

LEMMA C.1. *Consider a pure multitasking environment in which $G^j(x^j|a_j)$ satisfies both MLRP and NISP for some $j = 1, \dots, m$. Then $\mu_j > 0$ for any $a^* \in \text{int}(A)$.*

PROOF. Consider the agent's incentive to deviate marginally from some interior a^*_j,

$$EU_j(\mathbf{a}^*) = -\int \cdots \int \left(\int [-v(w(\mathbf{x}))] g^j_{a_j}(\mathbf{x}^j|a^*_j)\, d\mathbf{x}^j \right) \prod_{k \neq j} (g^k(\mathbf{x}^k|a^*_k)\, d\mathbf{x}^k) - c_j(\mathbf{a}^*).$$

If $\mu_j < 0$, the composite function $[-v(w(\mathbf{x}))]$ is monotonically increasing in $\mathbf{x}^j$, by MLRP. Hence, by NISP, the inner integral is positive. Thus, $EU_j(\mathbf{a}^*) < 0$, which contradicts L-IC$_{\mathbf{a}^*}$. The same conclusion holds if $\mu_j = 0$ since $c_j > 0$. This concludes the proof. □

Assume now that the conditions in Lemma C.1 holds *for all* $j = 1, 2, \dots, m$. This assumption can be shown to be equivalent to the assumption that $F(\mathbf{x}|\mathbf{a})$ satisfies MLRP and NISP. Thus, for brevity, the statement that $F(\mathbf{x}|\mathbf{a})$ satisfies MLRP and NISP signifies that $G^j(\mathbf{x}^j|a_j)$ satisfies MLRP and NISP for all $j = 1, \dots, m$. Under these assumptions, $v(w(\mathbf{x}))$ as defined in (C.1) is monotonically increasing in all signals.

Likewise, by Lemma 2, $v(w(\mathbf{x}))$ is Δ-antitone in $\mathbf{x}$ if ω is n-antitone and $l^j_{a_j}(\mathbf{x}^j|a_j)$ is Δ-antitone in $\mathbf{x}^j$ *for all* $j = 1, 2, \dots, m$. The "for all" qualifier is again somewhat cumbersome. Hence, define

$$l_\nabla(\mathbf{x}|\mathbf{a}) = (l^1_{a_1}(\mathbf{x}^1|a_1), l^2_{a_2}(\mathbf{x}^2|a_2), \dots, l^m_{a_m}(\mathbf{x}^m|a_m))$$

as the gradient (with respect to $\mathbf{a}$) to $\ln f(\mathbf{x}|\mathbf{a})$, for fixed $\mathbf{x}$. Thus, $l_\nabla(\mathbf{x}|\mathbf{a})$ is a natural generalization of the likelihood ratio in the $m = 1$ case. Here, $l_\nabla(\mathbf{x}|\mathbf{a})$ is introduced simply for notational convenience and in particular to permit a more succinct statement of results. Thus, I say that $l_\nabla(\mathbf{x}|\mathbf{a})$ is Δ-antitone in $\mathbf{x}$ if each element in $l_\nabla(\mathbf{x}|\mathbf{a})$ is Δ-antitone in $\mathbf{x}$. That is, $l_\nabla(\mathbf{x}|\mathbf{a})$ is Δ-antitone in $\mathbf{x}$ if and only if $l^j_{a_j}(\mathbf{x}^j|a_j)$ is Δ-antitone in $\mathbf{x}^j$ for all j. Similarly, $l_\nabla(\mathbf{x}|\mathbf{a})$ is Δ^2-antitone in $\mathbf{x}$ if and only if $l^j_{a_j}(\mathbf{x}^j|a_j)$ is Δ^2-antitone in $\mathbf{x}^j$ for all j. Note that $F(\mathbf{x}|\mathbf{a})$ satisfies MLRP if $l_\nabla(\mathbf{x}|\mathbf{a})$ is Δ-antitone or Δ^2-antitone in $\mathbf{x}$.

In summary, $F(\mathbf{x}|\mathbf{a})$ inherits MLRP, NISP, and a Δ-antitone or Δ^2-antitone likelihood ratio from its constituent parts, the distributions $G^j(\mathbf{x}^j|a_j)$. The same cannot be said for CISP, LOCC, or CLOCC. In fact, it is necessary that $G^j(\mathbf{x}^j|a_j)$ satisfies LOCC ($G^j(\mathbf{x}^j|a_j)$ is convex in a_j) for all j in order for $F(\mathbf{x}|\mathbf{a})$ to satisfy LOCC ($F(\mathbf{x}|\mathbf{a})$ is convex in $\mathbf{a}$), but it is not sufficient. The same observation applies to CISP and CLOCC. I return to this issue momentarily. However, for now I simply assume that $F(\mathbf{x}|\mathbf{a})$ satisfies CISP, LOCC, or CLOCC. It is now trivial to derive counterparts to the justifications of the FOA presented in Section 4. The proof of the next proposition follows the standard logic.

PROPOSITION C.1. *Consider a pure multitasking environment. Assume the second-best action is in* int(A). *Then the FOA is valid if any one of the following statements holds:*

(i) The function $F(x|a)$ satisfies MLRP, NISP, and CISP.

(ii) The function $F(x|a)$ satisfies NISP and LOCC, $l_\nabla(x|a)$ is Δ-antitone in x, and ω is n-antitone.

(iii) The function $F(x|a)$ satisfies NISP and CLOCC, $l_\nabla(x|a)$ is Δ^2-antitone in x, and ω is $2n$-antitone.

Proposition C.1 is a natural extension of the results in Section 4. However, given the previous observation that $F(\mathbf{x}|\mathbf{a})$ inherits MLRP, NISP, and a Δ-antitone likelihood ratio from $G^j(\mathbf{x}^j|a_j)$, it is natural to ask what conditions on $G^j(\mathbf{x}^j|a_j)$ are required for $F(\mathbf{x}|\mathbf{a})$ to satisfy, e.g., LOCC. Such a question is also in the spirit of the additive property discussed in Section 5. The difference is that here additional tasks are added to the benchmark model, whereas extra signal are added in Section 5.

Recall that a function must be convex if it is log convex. Next, note that $F(\mathbf{x}|\mathbf{a})$ is log convex in $\mathbf{a}$ if and only if for all j, $G^j(\mathbf{x}^j|a_j)$ is log convex in a_j. Thus, a sufficient (but not necessary) condition for LOCC is that $G^j(\mathbf{x}^j|a_j)$ is log convex in a_j for all j. For similar reasons, CLOCC is satisfied if $\int_{\mathbf{y}^j\leq\mathbf{x}^j} G^j(\mathbf{y}^j|a_j)\,d\mathbf{y}^j$ is log convex.

PROPOSITION C.2. *In the pure multitasking model, LOCC is satisfied if $G^j(\mathbf{x}^j|a_j)$ is log convex in a_j for all j. CLOCC is satisfied if $\int_{\mathbf{y}^j\leq\mathbf{x}^j} G^j(\mathbf{y}^j|a_j)\,d\mathbf{y}^j$ is log convex in a_j for all j.*

Since log convexity is preserved under integration, the first condition implies the second. This is similar to how Rogerson's CDFC implies Jewitt's condition. Although log convexity may seem like a substantial strengthening of convexity, many common examples in fact have both properties. In the univariate case, the distribution $G(x|a)=x^a$, $x\in[0,1]$, $a>0$, is log convex. Rogerson uses this distribution as an example of one that satisfies both MLRP and CDFC. The two distributions in footnote 5 are also log convex. Incidentally, Ábrahám et al. (2011) show that in a two-period model with hidden borrowing, the FOA is valid under a set of assumptions that include log convexity. See also Kirkegaard (2015).

As mentioned, log convexity is not necessary. In the special case of the SQIT model, each $G^j(\mathbf{x}^j|a_j)$ need only satisfy CISP, LOCC, or CLOCC. To begin, recall that in the SQIT model a direct proof that $\mu_j>0$ is possible (Appendix B); NISP is not needed in this case. From (B.1), given the FOA contract satisfies L-IC$_{\mathbf{a}^*}$, the agent's utility from action $\mathbf{a}$ can be written as

$$EU(\mathbf{a})=\overline{u}+\sum_{j=1}^{m}\left(\frac{c_j(\mathbf{a}^*)}{V_j(a_j^*)}\int l^j_{a_j}(\mathbf{x}^j|a_j^*)g^j(\mathbf{x}^j|a_j)\,d\mathbf{x}^j-(a_j-a_j^*)c_j(\mathbf{a}^*)\right)+(c^L(\mathbf{a}|\mathbf{a}^*)-c(\mathbf{a})).$$

By L-IC$_{a^*}$, the jth term in the first parentheses has a stationary point at $a_j = a_j^*$. If $l^j_{a_j}(\mathbf{x}^j|a_j^*)$ and $G^j(\mathbf{x}^j|a_j)$ jointly satisfy any of the conditions in Section 4 (e.g., in Propositions 3 or 4), then that stationary point identifies a global maximum.[8] Assuming this is true for all j, the sum of these terms is maximized at $\mathbf{a} = \mathbf{a}^*$. Since $c(\mathbf{a})$ is convex, the last term is also maximized at $\mathbf{a} = \mathbf{a}^*$ (where it is zero). Thus, G-IC$_{a^*}$ is satisfied.

PROPOSITION C.3. *Consider the SQIT model. Assume the second-best action is in* $\text{int}(A)$. *Then the FOA is valid if for each task, $j = 1, 2, \ldots, m$, one of the following conditions is met (but not necessarily the same condition for all j):*

(i) The function $G^j(x^j|a_j)$ satisfies MLRP and CISP.

(ii) The function $G^j(x^j|a_j)$ satisfies LOCC and $l^j_{a_j}(x^j|a_j^*)$ is Δ-antitone in x^j.

(iii) The function $G^j(x^j|a_j)$ satisfies CLOCC and $l^j_{a_j}(x^j|a_j^*)$ is Δ^2-antitone in x^j.

Propositions C.1–C.3 are relevant for the problem of determining how many tasks to assign an agent to, as they uncover conditions under which the FOA remains valid as more tasks are added. Similarly, the FOA can be used to investigate how to allocate a fixed number of tasks among a set of agents. The next section examines the impact of multitasking on the optimal contract.

APPENDIX D: COMPARING THE SQIT AND LEN MODELS

This section demonstrates that the pure multitasking environment has different equilibrium properties and yields different predictions than the LEN model. The SQIT model is ideally suited for this purpose.

Thus, consider the general SQIT model. From the principal's point of view, (B.2) reveals that the marginal cost of inducing higher a_j is

$$K_j(\mathbf{a}) = \left(\frac{\overline{u} + c(\mathbf{a})}{2}\right) c_j(\mathbf{a}) + \left(\frac{c_j(\mathbf{a}) c_{jj}(\mathbf{a})}{2V_j(a_j)} - \frac{c_j(\mathbf{a})^2 V_j'(a_j)}{4V_j(a_j)^2}\right) + \sum_{k \neq j} \frac{c_k(\mathbf{a})}{2V_k(a_k)} c_{kj}(\mathbf{a}).$$

The first term equals the marginal costs under full information. Holding fixed the other tasks, the second term essentially coincides with the marginal agency costs in a single-task setting (examined in detail in Section 6). The last term captures the interaction between tasks. Recall from Appendix B that the L-IC multiplier on the kth task is $\mu_k = c_k(\mathbf{a})/(2V_k(a_k))$. Now the marginal cost of task k changes as a_j changes, thus changing the last term in the L-IC constraint on task k. The last term in $K_j(\mathbf{a})$ thus reflects the costs of this interdependence between the incentive compatibility constraints of different tasks. Naturally, this is a cost if tasks k and j are substitutes, or $c_{kj} > 0$. However, it is a beneficial effect if the tasks are complements, or $c_{kj} < 0$.

Consider two tasks, a_1 and a_2. Assume in the remainder that the agent's cost function, $c(\mathbf{a})$, is symmetric in the tasks. Imagine, for now, that $a_1 = a_2$, i.e., the agent works

[8] The "cost function" $(a_j - a_j^*)c_j(\mathbf{a}^*)$ is convex in a_j. Thus, the standard argument applies.

equally hard on both tasks. By symmetry, it then holds that $c_1(\mathbf{a}) = c_2(\mathbf{a})$, $c_{11}(\mathbf{a}) = c_{22}(\mathbf{a})$, and $c_{k1}(\mathbf{a}) = c_{k2}(\mathbf{a})$ for $k = 3, \ldots, m$. By convexity, it is moreover the case that $c_{11}(\mathbf{a}) = c_{22}(\mathbf{a}) \geq c_{12}(\mathbf{a})$ at such a point. In fact, at such a point,

$$K_1(\mathbf{a}) - K_2(\mathbf{a}) = \frac{1}{2}c_1(\mathbf{a})(c_{11}(\mathbf{a}) - c_{12}(\mathbf{a}))\left(\frac{1}{V_1(a_1)} - \frac{1}{V_2(a_2)}\right) - \frac{1}{4}c_1(\mathbf{a})^2\left(\frac{V_1'(a_1)}{V_1(a_1)^2} - \frac{V_2'(a_2)}{V_2(a_2)^2}\right). \tag{D.1}$$

It is instructive to examine a few special cases. Example 7 illustrates that the agent may be induced to work hardest on the least informative task. Example 8 demonstrates that even if two tasks are perfect substitutes, the agent may be induced to work on both. Neither property arises in the LEN model.

Example 7. Assume first that the last term in (D.1) is zero and that $c_{11}(\mathbf{a}) > c_{12}(\mathbf{a})$ at $a_1 = a_2$. Then, at $a_1 = a_2$, $K_1(\mathbf{a}) \leq K_2(\mathbf{a})$ if and only if $V_1(a_1) \geq V_2(a_2)$. In other words, the most informative task has the lowest marginal costs. This property also holds in the LEN model, which is why Holmström and Milgrom (1988) (assuming $B(\mathbf{a})$ is symmetric as well) find that, "[t]he activity which can be measured with less noise will be relatively favored." See also the Dixit (1997) quote in Section 7.

However, this property does not hold in the SQIT model more generally, as the last term in (D.1) may outweigh the first. For instance, consider the case where $c(\mathbf{a}) = c(a_1 + a_2 + \cdots + a_m)$, $a_i \geq 0$ for all $i = 1, \ldots, m$. Note that (D.1) is valid for all $\mathbf{a}$ in this case (not just when $a_1 = a_2$), as it is always true that $c_1(\mathbf{a}) = c_2(\mathbf{a})$ and $c_{11}(\mathbf{a}) = c_{22}(\mathbf{a}) = c_{12}(\mathbf{a})$. Using such a cost function, Holmström and Milgrom (1991, Section 3) interpret a_j as time or attention devoted to a specific activity. Here the first term in (D.1) is zero. For any $\mathbf{a}$, it is possible that task 1 will be the most informative, $V_1(a_1) > V_2(a_2)$, yet will have the higher marginal cost, $K_1(\mathbf{a}) > K_2(\mathbf{a})$. This occurs if

$$\frac{V_1'(a_1)}{V_1(a_1)^2} < \frac{V_2'(a_2)}{V_2(a_2)^2}. \tag{D.2}$$

This condition is satisfied if, e.g., $V_1'(a_1) < 0 < V_2'(a_2)$. Assume now that $V_1(a_1) > V_2(a_2)$ and that (D.2) holds whenever $a_2 \leq a_1$. Hence, $K_1(\mathbf{a}) > K_2(\mathbf{a})$ whenever $a_2 \leq a_1$. To facilitate comparison with Holmström and Milgrom (1988), assume that $B(\mathbf{a})$ is symmetric. For concreteness, assume $B(\mathbf{a}) = \sum_{i=1}^m b(a_i)$, where $b(\cdot)$ is strictly increasing and strictly concave. Assuming the optimum is interior and satisfies $a_2 \leq a_1$, the principal's first-order conditions now yield

$$b'(a_1) = K_1(\mathbf{a}) > K_2(\mathbf{a}) = b'(a_2).$$

However, this contradicts the strict concavity of $b(\cdot)$. Thus, the solution is either not interior (the agent is not induced to work on both tasks) or it satisfies $a_2 > a_1$. In the latter case, the agent thus works harder on the least informative task. This outcome reflects the fact that the value of improving the informativeness of a task is greater when the task is less informative to begin with. Finally, interior actions are optimal when $b'(0)$ is large enough.

EXAMPLE 8. Assume that $c(\mathbf{a}) = c(a_1 + a_2 + \cdots + a_m)$ and $B(\mathbf{a}) = b(a_1 + a_2 + \cdots + a_m)$, such that tasks are perfect substitutes for both parties. Then the LEN model predicts a corner solution: the agent will be induced to work harder than the absolute minimum only on the most informative task. Again, this result is not robust. That is, the principal may induce the agent to work on multiple tasks, even though tasks are perfect substitutes and have different levels of informativeness. To see this more formally, assume $m = 2$. Consider all (a_1, a_2) combinations that yields some target total effort level, $a^t = a_1 + a_2$. For any interior pair, the cost to the principal is

$$K(a_1, a_2) = \left(\frac{\overline{u} + c(a^t)}{2}\right)^2 + \frac{c'(a^t)^2}{4}\left[\frac{1}{V_1(a_1)} + \frac{1}{V_2(a_2)}\right],$$

where only the last term depends on how total effort is distributed among tasks. Assuming that $\underline{a}_1 = \underline{a}_2 = 0$, the cost of inducing positive effort on only one task is

$$K(a_1, a_2) = \left(\frac{\overline{u} + c(a^t)}{2}\right)^2 + \frac{c'(a^t)^2}{4}\left[\frac{1}{V_j(a^t)}\right],$$

where j is the task that is being induced, $a_j = a^t$. Assume without loss of generality that $V_1(a^t) \geq V_2(a^t)$. In this case, it is optimal to induce positive effort on both tasks if and only if

$$\frac{1}{V_1(a_1)} + \frac{1}{V_2(a_2)} < \frac{1}{V_1(a^t)}$$

for some (a_1, a_2) pair with $a_1 + a_2 = a^t$. It is optimal to induce only task 1 if task 1 is unambiguously the most informative, or $\min V_1(a_1) \geq \max V_2(a_2)$. This result echoes the prediction from the LEN model. In particular, the signals' informativeness is completely *independent* of (a_1, a_2) in the LEN model. Thus, generically one of the tasks is always strictly more informative than the other (the exception is the symmetric case where the two are equally informative). However, the SQIT model allows more complicated interactions among informativeness. Thus, assume both terms on the left hand side of the above inequality are strictly increasing and strictly convex. Holmström's (1979) example satisfies this assumption. Note the implication that $V_i(a_i)$ is strictly decreasing. Assume that $V_1(a) = V_2(a) + \varepsilon$, where $\varepsilon > 0$ is small. Then it remains optimal to induce only task 1 if a^t is small. However, if a^t is sufficiently large, then it is preferable to induce both tasks. After all, since V_1 is decreasing in a_1, the information system becomes less and less informative if it relies only on task 1 (i.e., $a_1 = a^t$) as a^t becomes large. Eventually, it becomes preferable to lower a_1, thereby increasing V_1, by shifting effort toward task 2 even though the information value of the latter is mediocre. Note that even if tasks are completely symmetric, $V_2(a) = V_1(a)$, the distribution of total effort among tasks (whether positive effort on one or both tasks is induced) depends on the scale of the operation, i.e., on the total effort being targeted.[9]

[9] In Holmström's (1979) example, $1/V_i(a_i) = a_i^2$ and it follows that $1/V_1(\frac{1}{2}a^t) + 1/V_2(\frac{1}{2}a^t) < 1/V_1(a^t)$ for all $a^t > 0$. Thus, it is always optimal to induce the agent to work on both tasks. This example is special because $V_i(a_i) \to \infty$ as $a_i \to 0$, since F_i becomes degenerate in the limit.

References

Ábrahám, Árpád, Sebastian Koehne, and Nicola Pavoni (2011), "On the first-order approach in principal-agent models with hidden borrowing and lending." *Journal of Economic Theory*, 146, 1331–1361. [7]

Banker, Rajiv D. and Srikant M. Datar (1989), "Sensitivity, precision, and linear aggregation of signals for performance evaluation." *Journal of Accounting Research*, 27, 21–39. [2]

Boyer, Marcel and Bernard Sinclair-Desgagné (2001), "Corporate governance in the presence of major technological risks." In *Frontiers of Environmental Economics* (Henk Folmer, H. Landis Gabel, Shelby Gerking, and Adam Rose, eds.), 272–292, Edward Elgar, Cheltenham, UK. [5]

Dixit, Avinash (1997), "Power of incentives in private versus public organizations." *American Economic Review*, 87, 378–382. [9]

Holmström, Bengt (1979), "Moral hazard and observability." *Bell Journal of Economics*, 10, 74–91. [2, 10]

Holmström, Bengt and Paul Milgrom (1988), "Common agency and exclusive dealing." Unpublished, MIT. [9]

Holmström, Bengt and Paul Milgrom (1991), "Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design." *Journal of Law, Economics, and Organization*, 7, 24–52. [9]

Jewitt, Ian (1988), "Justifying the first-order approach to principal–agent problems." *Econometrica*, 56, 1177–1190. [3]

Jewitt, Ian, Ohad Kadan, and Jeroen M. Swinkels (2008), "Moral hazard with bounded payments." *Journal of Economic Theory*, 143, 59–82. [2]

Kirkegaard, René (2015), "Moral hazard with private rewards." Unpublished, University of Guelph. [7]

LiCalzi, Marco and Sandrine Spaeter (2003), "Distributions for the first-order approach to principal–agent problems." *Economic Theory*, 21, 167–173. [3]

Milgrom, Paul R. and Robert J. Weber (1982), "A theory of auctions and competitive bidding." *Econometrica*, 50, 1089–1122. [1]

Müller, Alfred and Dietrich Stoyan (2002), *Comparison Methods for Stochastic Models and Risks*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., West Sussex, England. [1]

Rogerson, William P. (1985), "The first-order approach to principal–agent problems." *Econometrica*, 53, 1357–1367. [3, 6]

Sinclair-Desgagné, Bernard (1999), "How to restore higher-powered incentives in multi-task agencies." *Journal of Law, Economics, and Organization*, 15, 418–433. [5]

Co-editor George J. Mailath handled this manuscript.

Manuscript received 29 October, 2014; final version accepted 13 December, 2015; available online 14 December, 2015.