

Lenells, Jonatan; Stea, Diego; Foss, Nicolai J.

Article

Optimal contracting under adverse selection: The implications of mentalizing

Contemporary Economics

Provided in Cooperation with:

University of Finance and Management, Warsaw

Suggested Citation: Lenells, Jonatan; Stea, Diego; Foss, Nicolai J. (2015) : Optimal contracting under adverse selection: The implications of mentalizing, Contemporary Economics, ISSN 2084-0845, Vizja Press & IT, Warsaw, Vol. 9, Iss. 2, pp. 215-231, <https://doi.org/10.5709/ce.1897-9254.168>

This Version is available at:

<https://hdl.handle.net/10419/141906>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Primary submission: 28.10.2014 | Final acceptance: 06.05.2015

Optimal Contracting under Adverse Selection: The Implications of Mentalizing

Jonatan Lenells¹, Diego Stea², Nicolai J. Foss²

ABSTRACT

We study a model of adverse selection, hard and soft information, and mentalizing ability—the human capacity to represent others' intentions, knowledge, and beliefs. By allowing for a continuous range of different information types, as well as for different means of acquiring information, we develop a model that captures how principals differentially obtain information on agents. We show that principals that combine conventional data collection techniques with mentalizing benefit from a synergistic effect that impacts both the amount of information that is accessed and the overall cost of that information. This strategy affects the properties of the optimal contract, which grows closer to the first best. This research provides insights into the implications of mentalizing for agency theory.

KEY WORDS:

Adverse selection, mentalizing, hard information, soft information, contract

JEL Classification: D82; D83

¹Department of Mathematics, KTH Royal Institute of Technology, Stockholm, Sweden; ²Copenhagen Business School - Department of Strategic Management and Globalization, Frederiksberg, Denmark

1 Introduction

Agency theory posits that informational asymmetry, whether modeled as an instance of hidden action or hidden knowledge, hinders the contracting parties from obtaining the first-best outcome (Holmström, 1979; Laffont & Martimort, 2002; Ross, 1973). The theory also allows individuals to partly reduce informational barriers by (in the case of the agent) signaling or (in the case of the principal) learning the agent's type and monitoring his effort. These activities, however, are treated in a highly stylized manner. For instance, in the standard moral hazard model, all signals

on the agent's effort can be included in the contract between the principal and the agent and are assumed to be verifiable. In fact, many signals on agents' efforts are verifiable but some are not, and principals may rely on non-verifiable information (e.g., body language and facial expressions) to assess an agent's effort. Similarly, in the adverse selection model, principals may rely on such soft psychological information in assessing agents' types.

In other words, information is an essential component of agency theory, and yet it is often modeled in a way that abstracts from some potentially key features of the real world. This paper addresses exactly this problem. Specifically, information differs substantially depending on its form, and recent research has begun to capture this fact by classifying it in terms of how hard *versus* soft it is (Godbillon-Camus & Godlewski,

Correspondence concerning this article should be addressed to: **Nicolai J. Foss**, Copenhagen Business School - Department of Strategic Management and Globalization, Kilevej 14, Frederiksberg 2000, Denmark. E-mail: njf.smg@cbs.dk

2006; Peterson, 2004). Hard information (e.g., a person's education level, experience, or income) is easily reduced to numbers, it can be collected in an impersonal way, and its meaning is less contingent on subjective judgements, opinions, or perceptions. On the other hand, soft information (e.g., a person's feelings, perceptions, values, or motivations) is difficult to accurately reduce to a numeric score, and its meaning is highly dependent on the context in which it is collected and on the personal opinions and perceptions of the person collecting it.

If information differs in terms of how hard or soft it is, it is pertinent to ask whether there are ways of obtaining it that are particularly suitable, depending on the type of information. Recent convergent developments in evolutionary anthropology (Call & Tomasello, 2008), cognitive neuroscience (Gallagher & Frith, 2003), and neuroeconomics (Singer & Fehr, 2005) highlight the importance of players' mentalizing—that is, their intersubjective understanding of preferences, intentions, knowledge, and beliefs. Information about these mental states is soft in nature and is crucially important in making sense of and predicting the behaviors of others (Singer & Fehr, 2005). Thus, mentalizing is ideally suited for the acquisition of soft information.

There is no reason to suppose that principals should not make use of mentalizing as a preferred method of inferring information about other players. As Singer and Fehr (Singer & Fehr, 2005) note, however, economists take a technical shortcut by assuming a common prior distribution over agent types without considering the determinants of this distribution. In other words, agency theory does not make explicit room for mentalizing. Yet, the theory effectively if implicitly assumes that the principal has perfect access to and knowledge of certain mental states of the agent (Foss & Stea, 2014). For example, in the standard moral hazard model (Holmström, 1979), the principal is assumed to know the risk preferences and reservation utility of the agent.

The object of this paper is to provide insights into the implications of mentalizing for agency theory. We base our analysis on a manager-worker relationship under adverse selection (Laffont & Martimort, 2002) where we allow for a continuous range of different information types and different means of acquiring in-

formation. We obtain three main sets of results. First, we show that mentalizing can be a low-cost method of acquiring information. Second, we show that mentalizing provides access to information that may be difficult to elicit in other ways. Third, we highlight that mentalizing impacts the design of the bilateral contract that the principal and agent sign, resulting in an increase in the volume of trade achieved under asymmetric information. All in all, this research suggests that a more nuanced description of how principals differentially obtain information on agents leads to a more accurate modeling of agency relationships.

2 The basic model with an informative signal

The basic adverse selection model with an informative signal (Laffont and Martimort, 2002) includes a principal P and an agent A . The principal wants to delegate to the agent the production of q units of a good. The value for the principal of these q units is $S(q)$, where $S(q)$ is a strictly increasing concave function (i.e., $S'(q) > 0$ and $S''(q) < 0$ for all q) such that $S(0) = 0$. The cost for the agent to produce q units is $C(q, \theta) = \theta q$, where $\theta > 0$ is the type of the agent. In exchange for the production of the q units, the agent receives a transfer t from the principal. The agent's utility is

$$U = t - \theta q,$$

while the principal's utility is

$$V = S(q) - t.$$

If the principal offers the agent a transfer $t \in R$ in exchange for the production of $q > 0$ units, we say that the principal offers the agent a (q, t) contract.

For simplicity, we assume that the agent can be of only two types: He is either *efficient* $\theta = \underline{\theta}$ or *inefficient* $\theta = \bar{\theta}$, where $\underline{\theta} < \bar{\theta}$. The cost for the agent to produce q units is $\underline{\theta}q$ if he is efficient and $\bar{\theta}q$ if he is inefficient.

It is common knowledge that the probability that $\theta = \underline{\theta}$ is $v \in (0, 1)$, while the probability that $\theta = \bar{\theta}$ is $1 - v$. Before the contracting process begins the agent discovers his type, but the principal only receives a signal σ with certain probabilistic information about θ .

Thus, the agent has more information than the principal (the agent has *hidden knowledge*). This asymmetry in information is the reason that only a second-best solution can be achieved.

For simplicity, we assume that σ may take only two values, σ_1 and σ_2 . Let the conditional probabilities of these respective realizations of the signal be

$$\mu_1 = \Pr(\sigma = \sigma_1 | \theta = \underline{\theta}) \geq \frac{1}{2}$$

$$\text{and } \mu_2 = \Pr(\sigma = \sigma_2 | \theta = \bar{\theta}) \geq \frac{1}{2}.$$

If $\mu_1 = \mu_2 = 1/2$, the signal is uninformative. Otherwise, the signal σ_1 brings good news in the sense that it is more likely that the agent is efficient if $\sigma = \sigma_1$ than if $\sigma = \sigma_2$.

Let us consider the case where the principal offers a menu of contracts $\{(q, t), (\bar{q}, \bar{t})\}$ hoping that an agent of type $\underline{\theta}$ will select (q, t) and an agent of type \bar{t} will select (\bar{q}, \bar{t}) . The timing is as follows:

1. The agent discovers his type $\theta \in \{\underline{\theta}, \bar{\theta}\}$.
2. The principal receives the signal $\sigma \in \{\sigma_1, \sigma_2\}$.
3. The principal offers a menu of contracts.
4. The agent accepts one or none of the contracts.
5. If a contract is accepted, the contract is executed.

Before receiving the signal σ , the principal expects the agent to be efficient with probability ν . After receiving the signal σ , the principal can compute an updated probability that the agent is efficient. According to Bayes' law, after receiving the signal σ the principal expects that the agent is efficient with probability

$$\hat{\nu}_1 = \Pr(\theta = \underline{\theta} | \sigma = \sigma_1) = \frac{\nu\mu_1}{\nu\mu_1 + (1-\nu)(1-\mu_2)}$$

if $\sigma = \sigma_1$, (2.1)

$$\hat{\nu}_2 = \Pr(\theta = \underline{\theta} | \sigma = \sigma_2) = \frac{\nu(1-\mu_1)}{\nu(1-\mu_1) + (1-\nu)\mu_2}$$

if $\sigma = \sigma_2$. (2.2)

2.1 Optimal contracts

The requirement that agent $\underline{\theta}$ (resp. $\bar{\theta}$) weakly prefers the contract (q, t) (resp. (\bar{q}, \bar{t})) leads to the following *incentive constraints*:

$$\underline{t} - \underline{\theta}q \geq \bar{t} - \underline{\theta}\bar{q}, \tag{2.3}$$

$$\bar{t} - \bar{\theta}\bar{q} \geq \underline{t} - \bar{\theta}q. \tag{2.4}$$

Moreover, for a menu to be accepted, the following two *participation constraints* must be satisfied:

$$\underline{t} - \underline{\theta}q \geq 0, \tag{2.5}$$

$$\bar{t} - \bar{\theta}\bar{q} \geq 0. \tag{2.6}$$

The principal's problem consists of finding the solutions $\{(q_j^{SB}, t_j^{SB}), (\bar{q}_j^{SB}, \bar{t}_j^{SB})\}$, $j = 1, 2$, of the two optimization problems

$$\sup_{\{(q, t), (\bar{q}, \bar{t})\}} [\hat{\nu}_j(S(q) - t) + (1 - \hat{\nu}_j)(S(\bar{q}) - \bar{t})]$$

subject to (2.3)-(2.6), $j = 1, 2$,

where $j = 1$ if $\sigma = \sigma_1$ and $j = 2$ if $\sigma = \sigma_2$. The solutions are given on p. 43 of (Laffont and Martimort, 2002); the optimal contract $\{(q_j^{SB}, t_j^{SB}), (\bar{q}_j^{SB}, \bar{t}_j^{SB})\}$ that the principal should offer if he receives the information signal σ_j is characterized by

$$S'(q_j^{SB}) = \underline{\theta}, S'(\bar{q}_j^{SB}) = \bar{\theta} + \frac{\hat{\nu}_j}{1 - \hat{\nu}_j} \Delta\theta,$$

$$t_j^{SB} = \underline{\theta}q_j^{SB} + \Delta\theta\bar{q}_j^{SB}, \bar{t}_j^{SB} = \bar{\theta}\bar{q}_j^{SB}, \tag{2.7}$$

where $\Delta\theta = \bar{\theta} - \underline{\theta}$ and $j = 1, 2$. In particular, the inefficient agent's production levels \bar{q}_1^{SB} and \bar{q}_2^{SB} associated with the signals σ_1 and σ_2 , respectively, satisfy

$$\begin{cases} S'(\bar{q}_1^{SB}) = \bar{\theta} + \frac{\hat{\nu}_1}{1 - \hat{\nu}_1} \Delta\theta = \bar{\theta} + \frac{\nu\mu_1}{(1-\nu)(1-\mu_2)} \Delta\theta, \\ S'(\bar{q}_2^{SB}) = \bar{\theta} + \frac{\hat{\nu}_2}{1 - \hat{\nu}_2} \Delta\theta = \bar{\theta} + \frac{\nu(1-\mu_1)}{(1-\nu)\mu_2} \Delta\theta. \end{cases} \tag{2.8}$$

Thus, compared with the first-best contract $\{(q^*, t^*), (\bar{q}^*, \bar{t}^*)\}$ for which $S'(\bar{q}^*) = \bar{\theta}$, the optimal contract entails a downward distortion of the inefficient agent's production in the presence of imperfect information. Indeed, because $S'' < 0$, the inequality $S'(\bar{q}_j^{SB}) > S'(\bar{q}^*)$ implies that $\bar{q}_j^{SB} < \bar{q}^*$. Because

$$\frac{1 - \mu_1}{\mu_2} \leq 1 \leq \frac{\mu_1}{1 - \mu_2},$$

the downward shift is larger if $\sigma = \sigma_1$ than if $\sigma = \sigma_2$.

This shows that

$$\bar{q}_1^{SB} \leq \bar{q}^{SB} \leq \bar{q}_2^{SB} < \bar{q}^*, \quad \underline{q}_1^{SB} = \underline{q}^{SB} = \underline{q}_2^{SB} = \underline{q}^*, \quad (2.9)$$

where $\{(q^{SB}, t^{SB}), (\bar{q}^{SB}, \bar{t}^{SB})\}$ is the second-best contract offered in the absence of an informative signal.

2.2 The principal's expected utility

Thus far we have followed (Laffont & Martimort, 2002). We now want to change our viewpoint slightly and formulate the optimization problem in terms of the principal's overall expected utility. This provides a way for us to merge the optimization problems for $\sigma = \sigma_1$ and $\sigma = \sigma_2$ into one problem.

The principal's expected utility when offering a menu of contracts $\{(q_j, t_j), (\bar{q}_j, \bar{t}_j)\}_{j=1}^2$ fulfilling the incentive and participation constraints is

$$\begin{aligned} EV_{((q_j, t_j), (\bar{q}_j, \bar{t}_j))} &= \Pr(\sigma = \sigma_1, \theta = \underline{\theta})[S(q_1) - t_1] + \\ &+ \Pr(\sigma = \sigma_1, \theta = \bar{\theta})[S(\bar{q}_1) - \bar{t}_1] + \\ &+ \Pr(\sigma = \sigma_2, \theta = \underline{\theta})[S(q_2) - t_2] + \\ &+ \Pr(\sigma = \sigma_2, \theta = \bar{\theta})[S(\bar{q}_2) - \bar{t}_2]. \end{aligned}$$

The basic optimization problem that maximizes the principal's utility is therefore

$$\begin{aligned} \sup_{((q_j, t_j), (\bar{q}_j, \bar{t}_j))} EV_{((q_j, t_j), (\bar{q}_j, \bar{t}_j))} &= \sup_{((q_j, t_j), (\bar{q}_j, \bar{t}_j))} \{v\mu_1[S(q_1) - t_1] + \\ &+ (1-\nu)(1-\mu_2)[S(\bar{q}_1) - \bar{t}_1] + \nu(1-\mu_1)[S(q_2) - t_2] + \\ &+ (1-\nu)\mu_2[S(\bar{q}_2) - \bar{t}_2]\}, \end{aligned} \quad (2.10)$$

where the contracts are subject to (2.3)-(2.6). Writing the right-hand side of (2.10) in the form

$$\begin{aligned} &(\nu\mu_1 + (1-\nu)(1-\mu_2)) \sup_{((q_1, t_1), (\bar{q}_1, \bar{t}_1))} \{\hat{v}_1[S(q_1) - t_1] + \\ &+ (1-\hat{v}_1)[S(\bar{q}_1) - \bar{t}_1]\} + \\ &+ (\nu(1-\mu_1) + (1-\nu)\mu_2) \sup_{((q_2, t_2), (\bar{q}_2, \bar{t}_2))} \{\hat{v}_2[S(q_2) - t_2] + \\ &+ (1-\hat{v}_2)[S(\bar{q}_2) - \bar{t}_2]\}, \end{aligned}$$

we see that the solution is given by (2.7). It follows that the principal's expected utility EV when offering the optimal menu of contracts is

$$\begin{aligned} EV &= v\mu_1[S(q_1^{SB}) - \underline{\theta}q_1^{SB} - \Delta\theta\bar{q}_1^{SB}] + \\ &+ (1-\nu)(1-\mu_2)[S(\bar{q}_1^{SB}) - \bar{\theta}\bar{q}_1^{SB}] + \\ &+ \nu(1-\mu_1)[S(q_2^{SB}) - \underline{\theta}q_2^{SB} - \Delta\theta\bar{q}_2^{SB}] + \\ &+ (1-\nu)\mu_2[S(\bar{q}_2^{SB}) - \bar{\theta}\bar{q}_2^{SB}] = \\ &= v[S(\underline{q}^*) - \underline{\theta}\underline{q}^* - \mu_1\Delta\theta\bar{q}_1^{SB} - (1-\mu_1)\Delta\theta\bar{q}_2^{SB}] + \\ &+ (1-\nu)[(1-\mu_2)S(\bar{q}_1^{SB}) + \mu_2S(\bar{q}_2^{SB}) + \\ &- (1-\mu_2)\bar{\theta}\bar{q}_1^{SB} - \mu_2\bar{\theta}\bar{q}_2^{SB}]. \end{aligned} \quad (2.11)$$

3 More information is better

Intuitively, we expect it to be advantageous for the principal to have access to additional information about the agent. In this section, we prove that this is indeed the case within the framework of the basic model of Section 2.

For simplicity, we henceforth suppose that $\mu_1 = \mu_2 = \mu$, where μ is the *informativeness* of the signal. Then, in view of (2.8),

$$\bar{q}_1^{SB} = S_q^{-1}(h(\mu)), \quad \bar{q}_2^{SB} = S_q^{-1}(h(1-\mu)),$$

where the function $h(\mu)$ is defined by

$$h(\mu) = \bar{\theta} + \frac{v\mu}{(1-\nu)(1-\mu)} \Delta\theta$$

and the inverse S_q^{-1} of $S_q \equiv S'$ exists because of our assumption that $S_{qq} < 0$. Equation (2.11) implies that

$$EV(\mu) = f(\mu) + f(1-\mu), \quad (3.1)$$

where the function $f(\mu)$ is defined by

$$\begin{aligned} f(\mu) &= \frac{v}{2}[S(\underline{q}^*) - \underline{\theta}\underline{q}^*] - \nu\mu\Delta\theta S_q^{-1}(h(\mu)) + \\ &+ (1-\nu)(1-\mu)[S(S_q^{-1}(h(\mu))) - \bar{\theta}S_q^{-1}(h(\mu))]. \end{aligned} \quad (3.2)$$

The expected utility in the absence of an informative signal is obtained by setting $\mu = 1/2$:

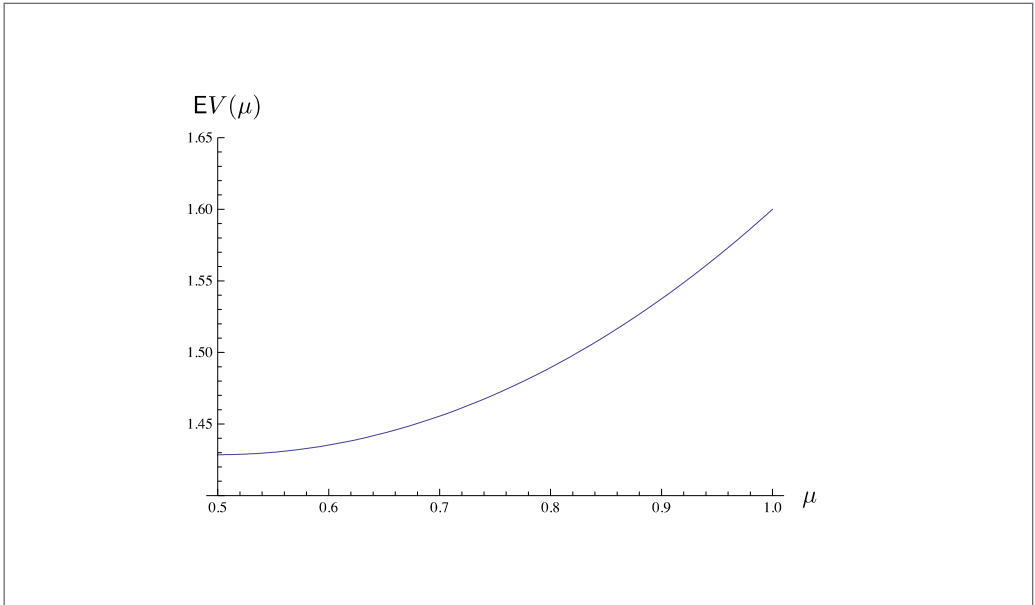


Figure 1. The graph of $EV(\mu)$ in the case of $S(q) = 2\sqrt{q}$ and the parameter values given in (3.4).

$$EV_{\text{nosignal}} = EV\left(\frac{1}{2}\right).$$

The following theorem expresses the fact that it is always beneficial for the principal to take additional information into account when formulating the contract. The more informative the signal σ is, the higher is the principal's expected utility.

Theorem 3.1 *The principal's expected utility function $EV(\mu)$ is strictly convex and attains its minimum at $\mu = 1/2$. In particular, $EV(\mu)$ is a strictly increasing function of μ for $\frac{1}{2} \leq \mu \leq 1$.*

Proof. We compute

$$\begin{aligned} f'(\mu) &= -v\Delta\theta S_q^{-1}(h(\mu)) + \\ &- (1-v)[S_q^{-1}(h(\mu)) - \bar{\theta} S_q^{-1}(h(\mu))] + \\ &- v\mu\Delta\theta \frac{d}{d\mu} S_q^{-1}(h(\mu)) + \\ &+ (1-v)(1-\mu)[S_q(S_q^{-1}(h(\mu))) - \bar{\theta}] \frac{d}{d\mu} S_q^{-1}(h(\mu)). \end{aligned} \quad (3.3)$$

The calculation

$$\begin{aligned} &-v\mu\Delta\theta + (1-v)(1-\mu)[S_q(S_q^{-1}(h(\mu))) - \bar{\theta}] = \\ &= -v\mu\Delta\theta + (1-v)(1-\mu) \frac{v\mu}{(1-v)(1-\mu)} \Delta\theta = 0 \end{aligned}$$

shows that the last two terms on the right-hand side of (3.3) cancel. Thus,

$$\begin{aligned} f'(\mu) &= -v\Delta\theta S_q^{-1}(h(\mu)) + \\ &- (1-v)[S_q^{-1}(h(\mu)) - \bar{\theta} S_q^{-1}(h(\mu))]. \end{aligned}$$

Differentiating once more, we find

$$\begin{aligned} f''(\mu) &= -v\Delta\theta \frac{d}{d\mu} S_q^{-1}(h(\mu)) + \\ &- (1-v)[S_q(S_q^{-1}(h(\mu))) - \bar{\theta}] \frac{d}{d\mu} S_q^{-1}(h(\mu)) = \\ &= [-v\Delta\theta - (1-v) \frac{v\mu}{(1-v)(1-\mu)} \Delta\theta] \frac{d}{d\mu} S_q^{-1}(h(\mu)) = \\ &= -\frac{v\Delta\theta}{1-\mu} \frac{d}{d\mu} S_q^{-1}(h(\mu)). \end{aligned}$$

Using that

$$\begin{aligned} \frac{d}{d\mu} S_q^{-1}(h(\mu)) &= \dots \\ &= \frac{h'(\mu)}{S_{qq}(S_q^{-1}(h(\mu)))} = \\ &= \frac{1}{S_{qq}(S_q^{-1}(h(\mu)))} \frac{v\Delta\theta}{(1-\nu)(1-\mu)^2} \end{aligned}$$

we obtain

$$f''(\mu) = -\frac{v^2(\Delta\theta)^2}{(1-\nu)(1-\mu)^3 S_{qq}(S_q^{-1}(h(\mu)))}$$

Because $S_{qq} < 0$ by assumption, this implies that

$$f''(\mu) > 0, \quad 0 < \mu < 1.$$

Hence,

$$(EV)''(\mu) = f''(\mu) + f''(1-\mu) > 0,$$

showing that $EV(\mu)$ is indeed strictly convex. Moreover, because $(EV)'(1/2) = f'(1/2) - f'(1/2) = 0$, $EV(\mu)$ attains its minimum at $\mu = 1/2$. This completes the proof.

Remark 3.2 The conclusion of Theorem 3.1 is reminiscent of the conclusion of Holmström's sufficiency theorem (Holmström, 1979). The contexts of these theorems differ in that the timing and setup of the contracting process are different.

Example 3.3 Consider the special case of $S(q) = 2\sqrt{q}$. In this case,

$$S_q(q) = q^{-1/2}, \quad S_q^{-1}(x) = x^{-2}, \quad S_{qq}(q) = -\frac{1}{2}q^{-3/2}.$$

Moreover, $EV(\mu)$ is given by (3.1)-(3.2) and

$$S_q^{-1}(h(\mu)) = \bar{q}_1^{SB} = [\bar{\theta} + \frac{v\mu}{(1-\nu)(1-\mu)} \Delta\theta]^{-2},$$

$$S_q^{-1}(\underline{\theta}) = \underline{q}^* = \underline{\theta}^{-2}.$$

Because

$$S_{qq}(S_q^{-1}(h(\mu))) = -\frac{1}{2}[\bar{\theta} + \frac{v\mu}{(1-\nu)(1-\mu)} \Delta\theta]^3,$$

we find

$$f'''(\mu) = \frac{2v^2(\Delta\theta)^2}{(1-\nu)(1-\mu)^3 \bar{\theta} + \frac{v\mu}{(1-\nu)(1-\mu)} \Delta\theta^3} > 0.$$

Hence, in accordance with Theorem 3.1, $(EV)''(\mu) = f''(\mu) + f''(1-\mu) > 0$. In Figure 1 the graph of $EV(\mu)$ is shown for the following choices of the parameters:

$$\bar{\theta} = 1, \quad \underline{\theta} = 0.5, \quad \nu = 0.6. \tag{3.4}$$

4 The basic model with a costly informative signal

We saw in the preceding section that the principal always benefits from additional information when formulating the contract. Thus, if information is free, the principal will always choose to acquire maximal information. In a more realistic scenario, there is a cost associated with the information in the signal σ (for example, the effort cost of the principal to obtain that information). In this section, we analyze the consequences of the information signal being costly.

Consider the model of Section 3 with an information signal of informativeness μ , where μ ranges from $1/2$ (no additional information) to 1 (full information), but suppose now that the information in the signal σ is costly for the principal. More precisely, suppose the principal's utility has the form

$$V = S(q) - t - C(\mu),$$

where $C(\mu)$ is the cost of obtaining a signal of informativeness μ . The principal's problem consists of solving the optimization problem (cf. (2.10))

$$\begin{aligned} \sup_{\{(q_j, t_j) \mid q_j, t_j\}} \\ \frac{1}{2} \leq \mu \leq 1} \{ &v\mu[S(\underline{q}_1) - t_1] + (1-\nu)(1-\mu)[S(\bar{q}_1) - \bar{t}_1] + \\ &+ \nu(1-\mu)[S(\underline{q}_2) - t_2] + (1-\nu)\mu[S(\bar{q}_2) - \bar{t}_2] - C(\mu) \} \end{aligned}$$

subject to (2.3)-(2.6).

For fixed μ , the solution is given by (2.7). The problem therefore reduces to maximizing the principal's expected utility

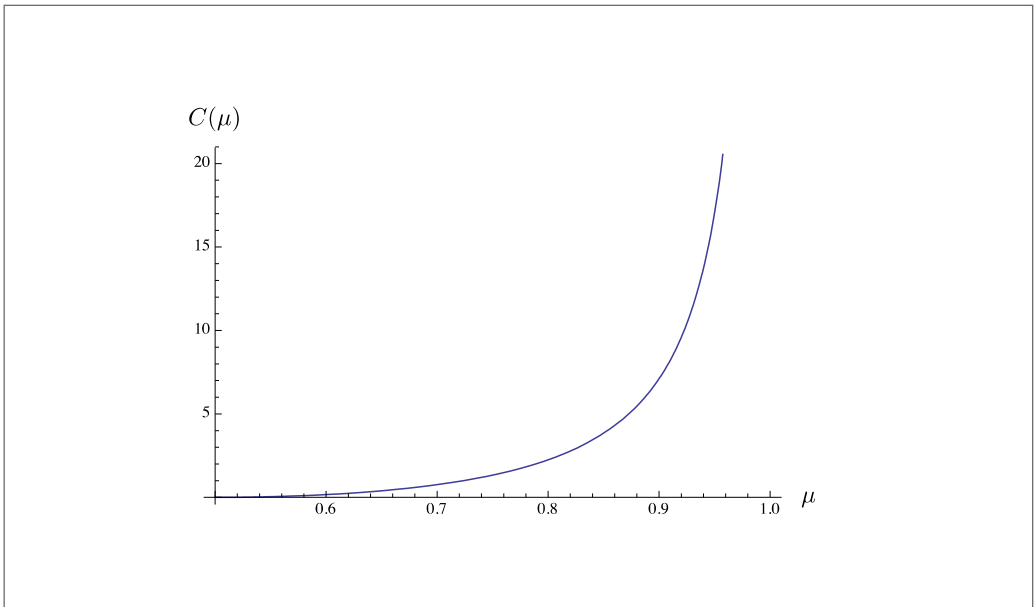


Figure 2. The graph of the function $C(\mu)$ in (4.2) for $c = 1$.

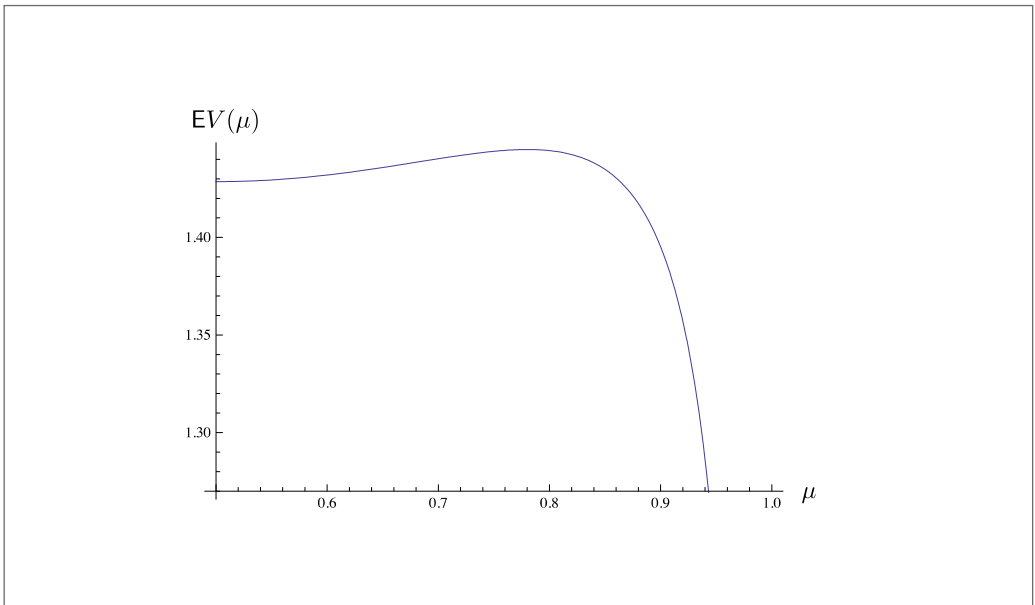


Figure 3. The graph of the principal's expected utility $EV(\mu)$ given by (4.1) as a function of μ with $S(q) = 2\sqrt{q}$, the cost function C given by (4.2), $c = 0.02$, and the parameter values given in (3.4).

$$EV(\mu) = f(\mu) + f(1-\mu) - C(\mu) \tag{4.1}$$

over $\mu \in [\frac{1}{2}, 1]$, where $f(\mu)$ is given by (3.2).

Example 4.1 Consider the information cost function $C(\mu)$ given by

$$C(\mu) = c\left(\frac{1}{\mu(1-\mu)} - 4\right), \quad \frac{1}{2} \leq \mu \leq 1. \tag{4.2}$$

where $c > 0$ is a constant, see Figure 2. Because $C'(\mu) > 0$ for $\mu \in (1/2, 1)$, a more informative signal is more costly. On the other hand, because $C(1/2) = 0$, an uninformative signal is free. For $S(q) = 2\sqrt{q}$, $c = 0.02$ and the parameter values in (3.4) the function $EV(\mu)$ in (4.1) is maximized for $\mu \approx 0.780$, see Figure 3. Thus, in this case, the optimal strategy for the principal is to invest a cost $C(0.780)$ in determining the agent's type before preparing the contract.

5 Different types of information

Information can assume substantially different forms, and recent research has begun to capture this fact by classifying information in terms of how hard (as opposed to soft) it is (Godbillon-Camus & Godlewski, 2006; Petersen, 2004). Hard information is either initially available in numbers or easy to reduce to numbers, can be collected in an impersonal way, and does not depend on the context of its production. For these reasons, hard information is rather standardized in nature and relatively easy to formalize and compare. For example, information about an agent's educational level or work experience is hard information. Soft information, on the other hand, is normally communicated in text and difficult to reduce to a numeric score. Its meaning is usually highly contingent on the context in which the information is collected and on the personal opinions and perceptions of the collector of the information. For these reasons, soft information is generally less standardized than hard information and more difficult to formalize and compare. For example, information about an agent's work motivation, preferences, or beliefs is soft information. Clearly, the extent to which a given piece of information can be hard or soft differs and the above distinction is to be seen as a continuum along which information can be classified (Petersen, 2004).

Given this, we now assume that there is a continuous range of different types of information. The in-

formation types are labeled by the variable $x \in [0, 1]$. Information of type $x = 0$ will be referred to as *hard* information whereas information of type $x = 1$ will be referred to as *soft* information. As x increases from 0 to 1 the information changes gradually from hard to soft. For each x , the principal can choose how much information $I(x)$ of type x to obtain, where $I(x)$ ranges from $I(x) = 0$ (he obtains no information of type x) to $I(x) = 1$ (he obtains maximal information of type x). We also assume that hard and soft information have the same weight in the overall informativeness of the signal. That is, equal amounts of hard and soft information provide the principal with equally useful insights on the agent.

Thus, assume that $\mu \in [1/2, 1]$ has the form

$$\mu = \frac{1}{2} \left(1 + \int_0^1 I(x) dx\right). \tag{5.1}$$

Because

$$0 \leq I(x) \leq 1, \quad x \in [0, 1], \tag{5.2}$$

the relation (5.1) implies that μ ranges from 1/2 to 1. We will denote by $C(x, y)$ the cost for the principal of obtaining an amount $y \in [0, 1]$ of information of type x .

The principal's problem consists of solving the optimization problem

$$\sup_{\substack{(q_j, t_j), (\bar{q}_j, \bar{t}_j) \\ I: [0, 1] \rightarrow [0, 1]}} \{ \nu \mu [S(\underline{q}_1) - t_1] + (1-\nu)(1-\mu)[S(\bar{q}_1) - \bar{t}_1] + \\ + \nu(1-\mu)[S(\underline{q}_2) - t_2] + (1-\nu)\mu[S(\bar{q}_2) - \bar{t}_2] - \int_0^1 C(x, I(x)) dx \}$$

subject to (2.3)-(2.6) and (5.1).

In view of (2.7), the problem reduces to maximizing the principal's expected utility

$$EV = f(\mu) + f(1-\mu) - \int_0^1 C(x, I(x)) dx. \tag{5.3}$$

over all functions $I: [0, 1] \rightarrow [0, 1]$ subject to (5.1). Assuming for simplicity that the constraint (5.2) is nowhere binding, variation with respect to $I(x)$ implies that

$$\frac{f'(\mu) - f'(1-\mu)}{2} \int_0^1 \delta I(x) dx - \int_0^1 \partial_x C(x, I(x)) \delta I(x) dx = 0$$

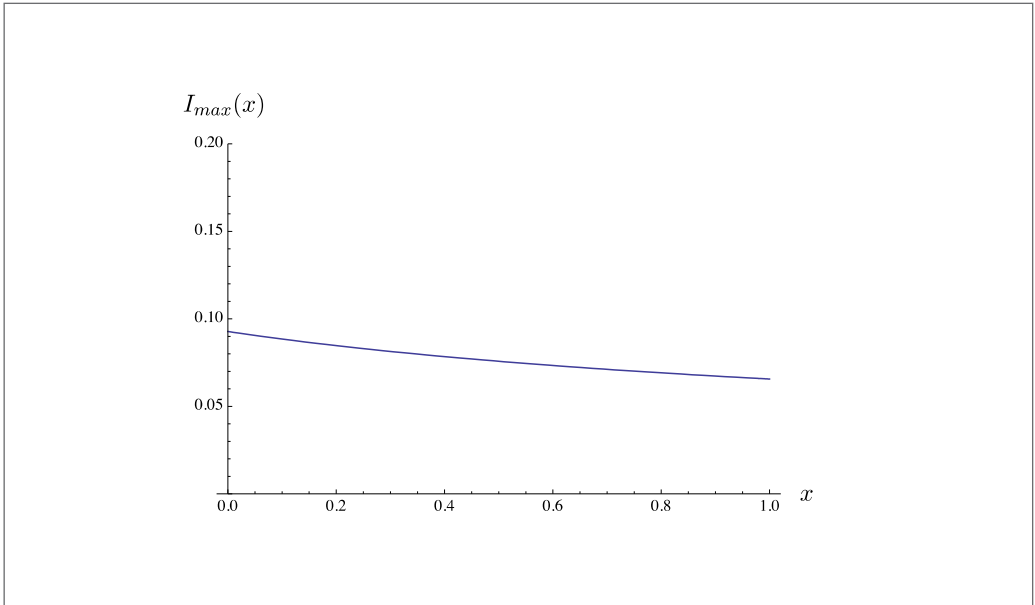


Figure 4. The graph of $I_{max}(x)$ for the example where $S(q) = 2\sqrt{q}$, $C(x, y) = (1+x)y^3$, and the parameter values in (3.4) have been used. As expected, the principal should acquire more hard than soft information.

for any variation $\delta I(x)$ around the maximum, where ∂_2 denotes partial differentiation with respect to the second argument. Thus, the optimal solution $I_{max}(x)$ is characterized by

$$f'(\mu_{max}) - f'(1 - \mu_{max}) = 2\partial_2 C(x, I_{max}(x)) \tag{5.4}$$

where

$$\mu_{max} = \frac{1}{2} (1 + \int_0^1 I_{max}(x) dx).$$

Example 5.1 Suppose that $S(q) = 2\sqrt{q}$ and $C(x, y) = (1+x)y^3$. This cost function models a scenario in which hard information (corresponding to $x = 0$) is cheaper than soft information (corresponding to $x = 1$). For a given type of information x , the marginal cost of information increases with the amount of information obtained (i.e., $\partial_2^2 C > 0$).

Let $F(\mu) := f'(\mu) - f'(1 - \mu)$. Because $\partial_2 C(x, y) = 3(1+x)y^2$, the condition (5.4) becomes

$$F(\mu_{max}) = 6(1+x)I_{max}(x)^2 \quad \text{i.e.} \quad I_{max}(x) = \sqrt{\frac{F(\mu_{max})}{6(1+x)}}. \tag{5.5}$$

This gives

$$\int_0^1 I_{max}(x) dx = \sqrt{F(\mu_{max})} \int_0^1 \frac{1}{\sqrt{6(1+x)}} dx = \sqrt{F(\mu_{max})} \frac{2(\sqrt{2}-1)}{\sqrt{6}}.$$

The value of μ_{max} is now determined by the condition (5.1):

$$\mu_{max} = \frac{1}{2} (1 + \sqrt{F(\mu_{max})} \frac{2(\sqrt{2}-1)}{\sqrt{6}}).$$

Once μ_{max} has been found by solving this equation, $I_{max}(x)$ is obtained from (5.5). This solves the optimization problem in the case when the constraint (5.2) is nowhere binding.

If we choose the parameter values in (3.4), we find $\mu_{max} \approx 0.5384$ and the function $I_{max}(x)$ takes the form shown in Figure 4. The constraint (5.2) is nowhere binding because $0 < I_{max}(x) < 1$ for all x . For these parameter values, the principal's expected

utility when acquiring information as prescribed by $I_{max}(x)$ is

$$EV_{max} = f(\mu_{max}) + f(1 - \mu_{max}) - \int_0^1 C(x, I_{max}(x)) dx \approx 1.4289. \quad (5.6)$$

Example 5.1 illustrates that when hard information is cheaper than soft information, the principal acquires more hard than soft information. Yet, why and under what conditions should hard information be assumed to be cheaper than soft information? In what follows, we address these questions by introducing the concept of mentalizing as a previously overlooked way of obtaining information.

6 Different ways of obtaining information

The ability to put oneself in another person's shoes has long been recognized as a crucial aspect of social interaction (Aumann & Brandenburger, 1995; Fudenberg & Tirole, 1991; Schutz, 1932; Weber, 1979). In particular, this ability serves as a key mechanism for coordinating beliefs and actions. Recent research in evolutionary anthropology (Call & Tomasello, 2008), cognitive neuroscience (Gallagher & Frith, 2003), and neuroeconomics (Singer & Fehr, 2005) highlights the importance of an individual's mentalizing processes—that is, her understanding of another individual's intentions, knowledge, and beliefs (Singer & Fehr, 2005). These are mental states that are not directly observable but are useful because they can make sense of, and predict, the behaviors of others (Call & Tomasello, 2008; Premack & Woodruff, 1978; Singer & Fehr, 2005).

First, the ability to understand *intentions*—that is, plans of action chosen in pursuit of a goal (Bratman, 1989; Dennet, 1987)—represents the first interpretive matrix for deciding what someone is doing (Tomasello et al., 2005). For example, suppose that an agent is working several extra hours and that a principal wants that agent to maintain his effort. The action of working extra hours, however, may have extremely different intentional connotations. The agent may be intrinsically motivated to deliver a good performance, or he may be externally motivated to do so by the potential for a monetary bonus. While giving a monetary reward to the extrinsically motivated agent could be a proper

way to incentivize him, giving the same reward to an intrinsically motivated agent may crowd out the intrinsic motivation and even diminish the overall effort (Frey & Jegen, 2001). Second, an agent's intentions are highly influenced by his *knowledge*. For this reason, the contextualization of an individual's intentions relative to an understanding of her knowledge is another fundamental constituent of mentalizing. In terms of the above example, if the principal knows that the agent knows that the organization has recently implemented a new reward system, the principal may expect the agent to work harder to obtain a bonus (rather than because of an innate interest in the task). Finally, as *beliefs* are by definition mental, the ability to understand someone's beliefs has been defined as the most complex component of mentalizing (Tomasello et al., 2005). In terms of the example, suppose that the principal knows that the agent works extra hours because of the reward system. Suppose the principal also knows that the agent is ignorant of the output-based (as opposed to input-based) nature of the reward criterion—in other words, the principal knows that the agent is wrong in thinking that his extra work will automatically result in higher compensation. The principal may benefit from this more nuanced understanding, and decide to not inform the agent about his mistaken belief.

Neuroscience research shows that humans have a brain system that is dedicated to mentalizing and that specific brain regions are activated when people engage in automatic as well as deliberate mentalizing (Frith & Frith 2003; Gallagher & Frith, 2003). Further, mentalizing may be understood as a skilled behavior in that it is program-like (i.e., mentalizing consists of an ordered sequence of cognitive steps), it is built upon a mixture of tacit and explicit knowledge (in fact, rarely is the mentalizer completely aware of the mechanisms that engender his having a theory of the other's mind), and it requires the making of a certain number of choices that vary in terms of the degree of intentionality (i.e., automatic *versus* deliberate mentalizing).

Information about an individual's intentions, knowledge, and beliefs is better captured in text than in a numeric score. Further, it must be collected in person and its meaning is likely to be highly contingent on the collector's opinions and perceptions. In other words, information about an agent's intentions, knowledge, and beliefs is soft in nature (Godbillon-Camus &

Godlewski, 2006; Petersen, 2004). Therefore, mentalizing is ideally suited for the acquisition of soft information. On the other hand, given its subtle psychological nature, mentalizing is not well suited for the acquisition of hard information, as this type of information is better captured by more conventional (hard) data collection techniques. The opposite applies to these other techniques, which are ideally suited for the acquisition of hard types of information, but not for that of soft information.

As an example, consider the following two pieces of information regarding a hypothetical software engineer (agent). First, the agent holds a master’s degree and has some years of work experience. Second, the agent loves the technical nature of his work and is entirely driven by this passion in his daily activities. A principal that would want to obtain the first bit of information (education and experience) would find it much easier to do so by simply looking at the hard data on the employee’s curriculum vitae. Clearly, he would find it extremely difficult to reach the same conclusions by exclusively mentalizing with the agent. On the other hand, there hardly is any way for a curriculum vitae to capture in an accurate and reliable way an agent’s innate passions and interests. Thus, a principal that wants to obtain this information would be better served by trying to put herself in the agent’s shoes—e.g., by looking at how the agent talks about his work-related activities—so as to have a feeling of what drives that agent in his function as software engineer. In other words, it is easier to obtain soft information via mentalizing than via hard data collection. Similarly, it is easier to obtain hard information via hard data collection than via mentalizing.

As in the preceding section, we assume that there exists a continuous range of different types of information labeled by the variable $x \in [0, 1]$, where $x = 0$ corresponds to hard information and $x = 1$ corresponds to soft information. In line with the above argumentation, we now also assume that the principal has two different ways of obtaining information: he can either use data collection or mentalizing. We will denote by $C_{dat}(x, y)$ and $C_{men}(x, y)$ the cost for the principal of obtaining an amount $y \in [0, 1]$ of information of the type x via data collection and mentalizing, respectively. Moreover, we let $I_{dat}(x)$ and $I_{men}(x)$ denote the amount of information of type x the principal obtains

via data collection and mentalizing, respectively. As in Section 5, we let $I(x)$ denote the total information obtained, that is,

$$0 \leq I_{dat}(x) + I_{men}(x) = I(x) \leq 1. \tag{6.1}$$

The principal’s problem consists of solving the optimization problem

$$\begin{aligned} & \sup_{\substack{(\underline{q}_j, \underline{t}_j), (\bar{q}_j, \bar{t}_j) \\ I_{dat} \in [0,1] \rightarrow [0,1] \\ I_{men} \in [0,1] \rightarrow [0,1]}} \{ \nu \mu [S(\underline{q}_1) - \underline{t}_1] + (1 - \nu)(1 - \mu) [S(\bar{q}_1) - \bar{t}_1] + \\ & + \nu(1 - \mu) [S(\underline{q}_2) - \underline{t}_2] + (1 - \nu)\mu [S(\bar{q}_2) - \bar{t}_2] + \\ & - \int_0^1 [C_{dat}(x, I_{dat}(x)) + C_{men}(x, I_{men}(x))] dx \} \end{aligned}$$

subject to (2.3)-(2.6), (5.1), and (6.1).

For a fixed choice of $I_{dat}(x)$ and $I_{men}(x)$, the solution is given by (2.7). The problem therefore reduces to

$$\begin{aligned} & \sup_{\substack{I_{dat} \in [0,1] \rightarrow [0,1] \\ I_{men} \in [0,1] \rightarrow [0,1]}} \{ f(\mu) + f(1 - \mu) + \\ & - \int_0^1 [C_{dat}(x, I_{dat}(x)) + C_{men}(x, I_{men}(x))] dx \} \end{aligned}$$

subject to (5.1) and (6.1).

Changing variables from $\{I_{dat}, I_{men}\}$ to $\{I_{dat}, I\}$, we can write this as

$$\begin{aligned} & \sup_{\substack{I_{dat} \in [0,1] \rightarrow [0,1] \\ I \in [0,1] \rightarrow [0,1]}} \{ f(\mu) + f(1 - \mu) + \\ & - \int_0^1 [C_{dat}(x, I_{dat}(x)) + C_{men}(x, I(x) - I_{dat}(x))] dx \} \end{aligned}$$

subject to (5.1) and $0 \leq I_{dat}(x) \leq I(x)$.

Because I_{dat} only appears in the integrand and the integrand is positive for every x , we can rewrite this as

$$\sup_{I \in [0,1] \rightarrow [0,1]} \{ f(\mu) + f(1 - \mu) - \int_0^1 C_{eff}(x, I(x)) dx \}$$

where $\mu = \frac{1}{2} (1 + \int_0^1 I(x) dx)$,

and the *effective cost function* $C_{eff}(x, y)$ is defined by

$$C_{eff}(x, y) = \min_{0 \leq z \leq y} [C_{dat}(x, z) + C_{men}(x, y - z)]. \tag{6.2}$$

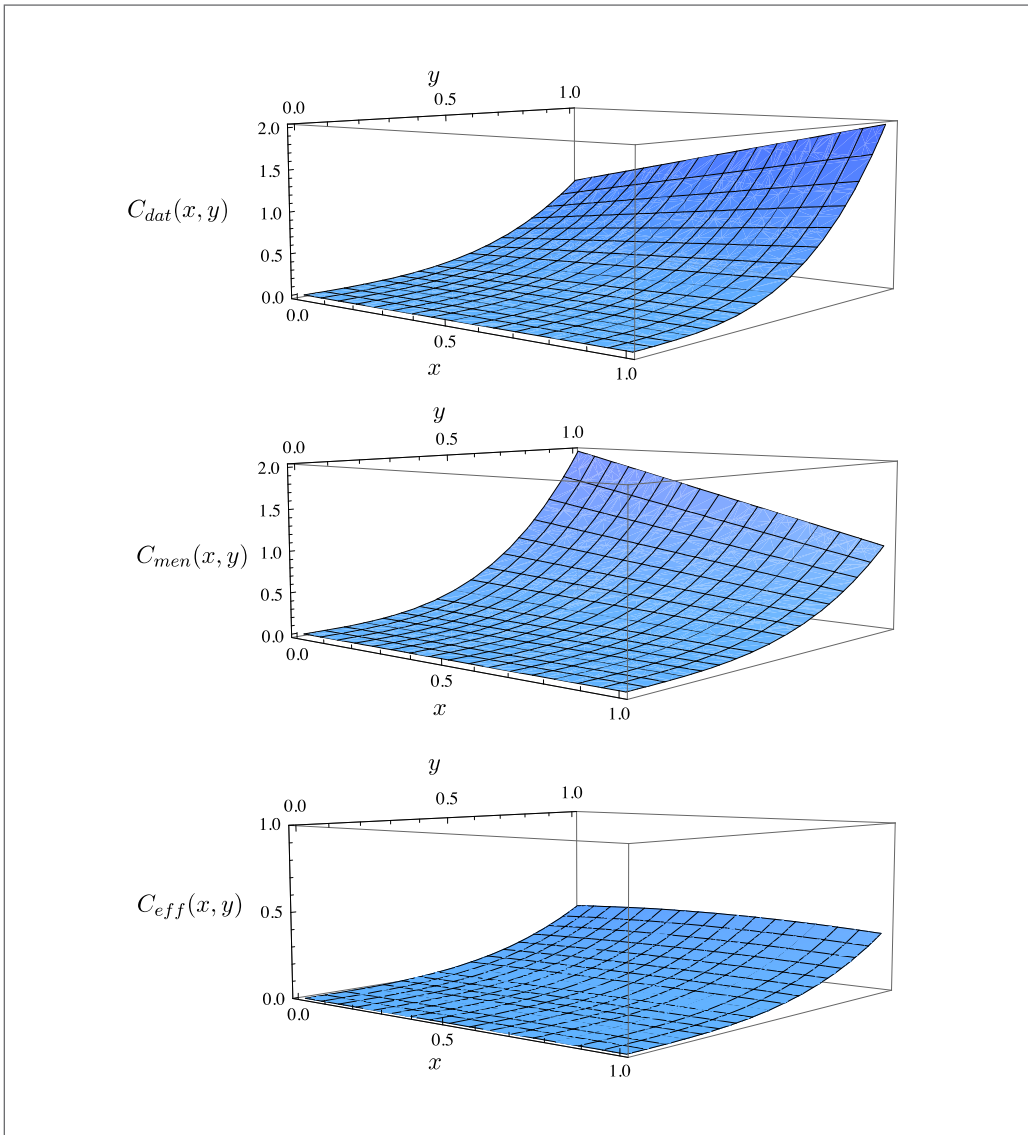


Figure 5. The graphs of the cost functions $C_{dat}(x, y)$ and $C_{men}(x, y)$ given in (6.3) and the corresponding effective cost function $C_{eff}(x, y)$ defined in (6.2).

This optimization problem is of the form considered in Section 5 and the optimal solution $I_{max}(x)$ is characterized by equation (5.4). Once $I_{max}(x)$ has been determined, $I_{dat,max}(x)$ is found from (6.2) according to

$$I_{dat,max}(x) = \underset{0 \leq z \leq I_{max}(x)}{\operatorname{argmin}} [C_{dat}(x, z) + C_{men}(x, I_{max}(x) - z)].$$

Example 6.1 Suppose that $S(q) = 2\sqrt{q}$ and

$$C_{dat}(x, y) = C_{men}(1 - x, y) = (1 + x)y^3. \tag{6.3}$$

In line with our previous discussion, these cost functions have the property that it is cheaper to obtain hard information via data collection than via mentalizing,

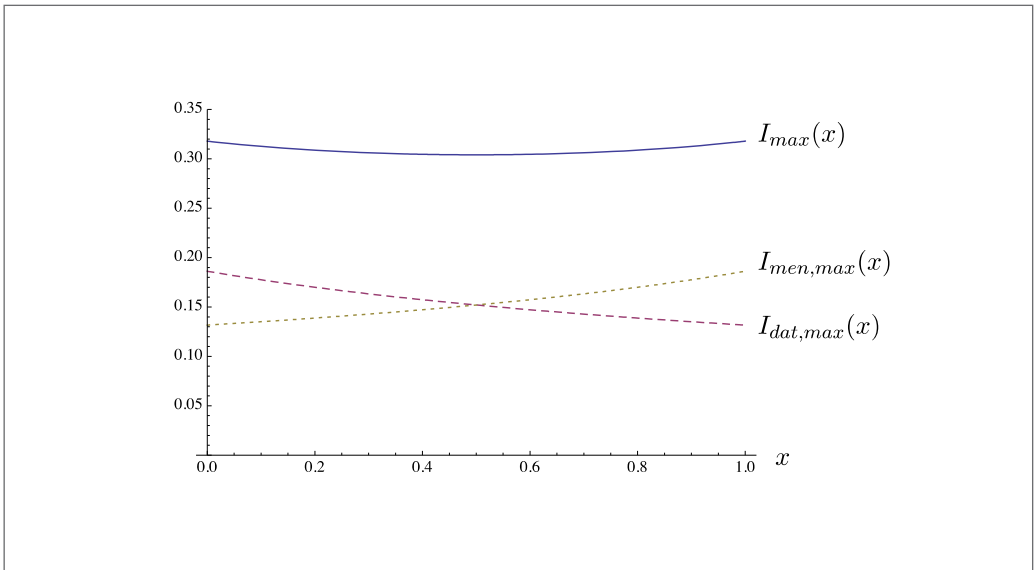


Figure 6. The graphs of the functions $I_{max}(x)$, $I_{dat,max}(x)$, and $I_{men,max}(x)$ given in (6.5) and (6.6). The graphs show that the principal should use primarily data collection (resp. mentalizing) to acquire hard (resp. soft) information.

while it is cheaper to obtain soft information via mentalizing than via data collection. Solving the condition

$$\frac{\partial}{\partial z}[C_{dat}(x, z) + C_{men}(x, y - z)] = 0$$

for z , we find that $C_{dat}(x, z) + C_{dat}(x, y - z)$ attains its minimum for $z = z_{min}(x, y)$ where

$$z_{min}(x, y) = \frac{(x - 2 + \sqrt{2 + x - x^2})y}{2x - 1} \in (0, 1). \tag{6.4}$$

Thus, the effective cost function defined in (6.2) is given by

$$C_{eff}(x, y) = \frac{(2 + x - x^2)(3 - 2\sqrt{2 + x - x^2})y^3}{(1 - 2x)^2}.$$

The graph of C_{eff} is displayed in Figure 5 and illustrates the synergistic effect of data collection and mentalizing on the cost of information. The simultaneous use of both data collection and mentalizing has an impact on the overall cost of accessing that information such that the average cost for any given amount of information is lower when the principal can selectively decide to use mentalizing or data collection depending on the nature of the information itself.

Letting $F(\mu) = f'(\mu) - f'(1 - \mu)$, the condition (5.4) yields

$$I_{max}(x) = \frac{\sqrt{F(\mu_{max})} |1 - 2x|}{\sqrt{6} \sqrt{(2 + x - x^2)(3 - 2\sqrt{2 + x - x^2})}}. \tag{6.5}$$

Thus

$$\int_0^1 I_{max}(x) dx = \frac{4(\sqrt{2} - 1)\sqrt{F(\mu_{max})}}{\sqrt{6}}.$$

The value of μ_{max} is now determined by the condition (5.1) which reads

$$\mu_{max} = \frac{1}{2} \left(1 + \frac{4(\sqrt{2} - 1)\sqrt{F(\mu_{max})}}{\sqrt{6}} \right).$$

For the parameter values in (3.4), we find $\mu_{max} \approx 0.6543$. Substituting this into (6.5), we find that the function $I_{max}(x)$ takes the form shown in Figure 6. It is clear from this figure that $0 < I_{max}(x) < 1$ for all x , ensuring that the constraint (5.2) is nowhere binding. In view of (6.4), the functions $I_{dat,max}(x)$ and $I_{men,max}(x)$ are given by

$$I_{dat,max}(x) = \frac{(x - 2 + \sqrt{2 + x - x^2})I_{max}(x)}{2x - 1},$$

$$I_{men,max}(x) = I_{max}(x) - I_{dat,max}(x). \tag{6}$$

Lastly, the principal's expected utility when acquiring information as dictated by $I_{max}(x)$ is

$$EV_{max} = f(\mu_{max}) + f(1 - \mu_{max}) - \int_0^1 C(x, I_{max}(x)) dx \approx 1.4339.$$

6.1 Comparisons

If the principal were to obtain the information described by the function $I_{max}(x)$ in (6.5) using only data collection (and did not have access to mentalizing), then the expected utility would decrease from 1.4339 to

$$EV_{max} = f(\mu_{max}) + f(1 - \mu_{max}) - \int_0^1 C_{dat}(x, I_{max}(x)) dx \approx 1.4005,$$

reflecting the fact that some information is cheaper to obtain via mentalizing than data collection. On the other hand, if the principal knew from the very beginning that he could only obtain information via data collection, he would conclude that his utility is maximized by acquiring information as dictated by the function $I_{max}(x)$ obtained in Example 5.1. Then, the expected utility would be 1.4289 (see Eq. (5.6)), which is again less than the expected utility 1.4339 obtained in the case when both data collection and mentalizing are available. This is despite the fact that the average costs of data collection and mentalizing coincide in the above example.

Similarly, the information obtained when the principal can only use data collection ($\mu_{max} \approx 0.5384$, Example 5.1) is significantly less than the information obtained when both data collection and mentalizing are used ($\mu_{max} \approx 0.6543$, Example 6.1).

In summary, by combining both data collection and mentalizing, the principal takes advantage of a synergistic effect which impacts both the amount of information that can be accessed and the overall cost of accessing that information. Mentalizing provides access to information that would not be possible to elicit using data collection only, and data collection provides access to information that would not be possible to elicit by using mentalizing only. Additionally, the average cost for any given amount of information is lower when the principal can selectively decide to use mentalizing or data collection depending on the nature of the information itself. This is reflected in the principal's expected utility,

which is higher when he can make use of both data collection and mentalizing.

7 Implications

What are the implications of mentalizing for the contracting process? We now explore how transfer and production levels, as well as the principal's expected utility, are impacted by the principal's mentalizing. We do so by means of a simple example.

Let the variable $m \geq 0$ denote the principal's degree of mentalizing ability, where $m = 0$ corresponds to no mentalizing skills whereas a large value of m corresponds to a high ability to mentalize. Additionally, let $S(q) = 2\sqrt{q}$ and consider the cost functions C_{dat} and C_{men} given by

$$C_{dat}(x, y) = e^x y^3, \quad C_{men}(x, y) = \frac{e^{1-x} y^3}{m}. \tag{7.1}$$

The parameter m is introduced in such a way that a high mentalizing ability m corresponds to a low cost of obtaining information via mentalizing. Proceeding as in Example 6.1, we can solve the optimization problem associated with (7.1). This determines the information acquisition functions $I_{dat,max}(x)$ and $I_{men,max}(x)$ as well as the associated menu of contracts $\{(q_j, t_j), (\bar{q}_j, \bar{t}_j)\}_{j=1}^2$ that maximize the principal's expected utility for each value of m . We then compute the principal's expected transfer to the efficient agent

$$E t_1^{SB} = \Pr(\sigma = \sigma_1 | \theta = \underline{\theta}) t_1^{SB} + \Pr(\sigma = \sigma_2 | \theta = \underline{\theta}) t_2^{SB} = \mu t_1^{SB} + (1 - \mu) t_2^{SB},$$

as well as the efficient agent's expected production level $E q_1^{SB} = \mu q_1^{SB} + (1 - \mu) q_2^{SB}$. For the inefficient agent, the analogous quantities are

$$(E \bar{q}^{SB}, E \bar{t}^{SB}) = ((1 - \mu) \bar{q}_1^{SB} + \mu \bar{q}_2^{SB}, (1 - \mu) \bar{t}_1^{SB} + \mu \bar{t}_2^{SB}).$$

In the end, we compare these second-best values with the first-best contract $\{(q^*, t^*), (\bar{q}^*, \bar{t}^*)\}$ characterized by

$$S'(q^*) = \underline{\theta}, \quad S'(\bar{q}^*) = \bar{\theta}, \quad t^* = \underline{\theta} q^*, \quad \bar{t}^* = \bar{\theta} \bar{q}^*.$$

For the parameter values specified in (3.4) we have $(q^*, t^*) = (4, 2)$ and $(\bar{q}^*, \bar{t}^*) = (1, 1)$. The expected second-best contracts as functions of m are displayed in

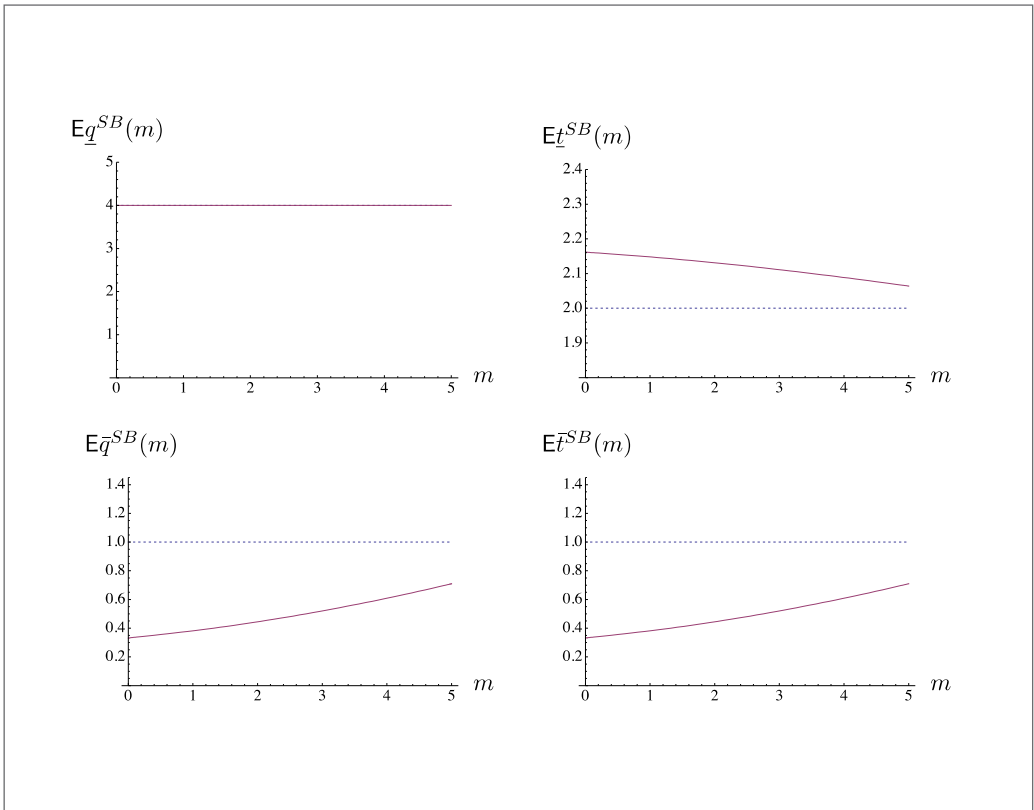


Figure 7. The expected transfers and production levels as functions of the mentalizing ability m (solid) and the corresponding first-best values (dotted). In the case of $E q^{SB}(m)$, the solid and dotted lines coincide because the efficient agent’s production level is the same for the first-best and second-best solutions, see equation (2.9).

Figure 7. We see from the figure that the second-best values approach the first-best values as the mentalizing ability m increases—that is, the information rent associated with imperfect information decreases with an increasing m . As m increases, the cost of retrieving information by means of mentalizing decreases. Hence, the principal can afford to acquire more information before designing the appropriate menu of contracts. The principal’s overall utility goes up accordingly and ultimately approaches the expected utility of the first-best scenario, which is given by (see Figure 8)

$$EV^* = \nu(S(\underline{q}^*) - \underline{t}^*) + (1 - \nu)(S(\bar{q}^*) - \bar{t}^*) = 1.6.$$

In summary, as Figures 7 and 8 illustrate, a high ability to mentalize drives the transfer and production levels,

and the utility function of the principal, closer to the first-best values that would be obtained in a scenario with perfect information.

Remark 7.1 As m increases beyond $m = 5$, the graphs of the second-best quantities displayed in Figures 7 and 8 flatten out and approach the corresponding first-best values. Indeed, for values of m larger than $m \approx 6.3$, the constraint $I(x) \leq 1$ saturates. The constraint first saturates for x near 1 and as m increases further, it saturates for smaller and smaller values of x . For very large (unrealistically large) values of m , the principal can afford acquiring full information so that $I_{max}(x) = 1$ for $0 \leq x \leq 1$. In this case, the principal can rely on perfect information when designing the contract and the first-best scenario is recovered.

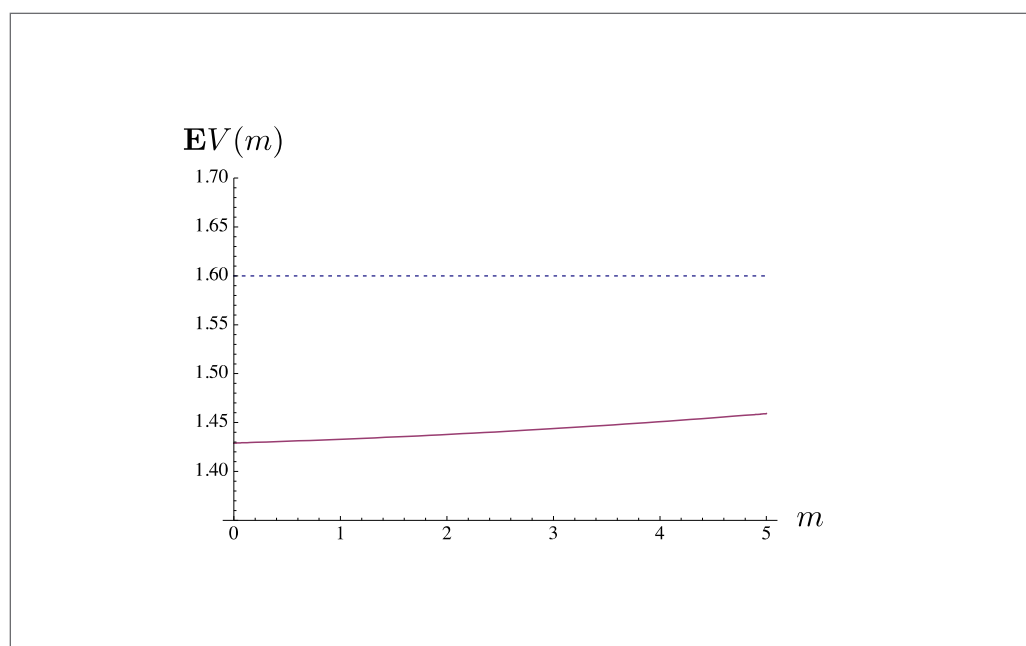


Figure 8. The principal's expected utility as a function of the mentalizing ability m (solid) and the corresponding first-best value (dotted).

8 Summary and conclusions

Recent developments in evolutionary anthropology (Call & Tomasello, 2008), cognitive neuroscience (Gallagher & Frith, 2003), and neuroeconomics (Singer & Fehr, 2005) highlight the importance of players' intersubjective understanding of preferences, intentions, and beliefs. When a player makes inferences about such mental states, she mentalizes—that is, she forms conjectures about mental states that are not directly observable but are useful because they can make sense of and predict the behaviors of others (Singer & Fehr, 2005).

The purpose of the present paper has been to define a space for mentalizing in principal-agent theory. We have taken some initial steps towards a more nuanced description of how principals differentially obtain information on agents by considering an extension of the basic adverse selection model, allowing for a continuous range of different information types as well as for different means of acquiring information. Our point of departure has been that principals are likely

to resort to both conventional data collection tools as well as mentalizing processes when extracting information about an agent. Given its subtle psychological nature, mentalizing is ideally suited for the processing of soft information. On the other hand, hard information is better captured by more conventional approaches.

Within the context of our model, we have observed that by combining both data collection and mentalizing, the principal can take advantage of a synergistic effect that impacts both the amount of information that can be accessed and the overall cost of accessing that information. This observation is reflected in the transfer and production levels as well as in the principal's utility. A high ability to mentalize drives these quantities closer to the first-best values that would be obtained in a scenario with perfect information. All in all, this research shows that a diversified approach to information acquisition leads to a refined preparation of the menu of contracts and a more efficient delegation.

References

- Aumann, R. J., Brandenburger, A. (1995). Epistemic Conditions for Nash Equilibrium. *Econometrica*, 63 (5), 1161–1180.
- Bratman, M. E. (1989). Intention and Personal Policies. Philosophical Perspectives. In J. E. Tomberlin (Ed.), *Philosophy of Mind and Action Theory*. Blackwell (vol. 3). Blackwell.
- Call, J., Tomasello, M. (2008). Does the Chimpanzee Have a Theory of Mind? 30 Years Later. *Trends in Cognitive Sciences*, 12 (5), 187–192.
- Dennet, D. C. (1978). *The intentional stance*. MIT Press.
- Frey, B. S., Jegen, R. (2001). Motivation Crowding Theory. *Journal of Economic Surveys*, 15 (5), 589–611.
- Frith, U., Frith, C. D. (2003). Development and Neurophysiology of Mentalizing. *Philosophical Transactions of the Royal Society of London, B: Biological Sciences*, 358, 459–473.
- Foss, N., Stea, D. (2014). Putting a realistic theory of mind into agency theory: Implications for reward design and management in principal-agent relations. *European Management Review*, 11 (1), 101–116.
- Fudenberg, D., Tirole, J. (1991). *Game Theory*. Cambridge, MA: MIT Press.
- Gallagher, H. L., Frith, C. D. (2003). Functional Imaging of “Theory of Mind”. *Trends in Cognitive Sciences*, 7 (2), 77–83.
- Godbillon-Camus, B., Godlewski, C. J. (2006). Credit Risk Management in Banks: Hard Information, Soft Information and Manipulation.
- Holmström, B. (1979). Moral hazard and observability. *Bell Journal of Economics*, 10 (1), 74–91.
- Laffont, J., Martimort, D. (2002). *The Theory of Incentives: The Principal-Agent Model*. Princeton, NJ: Princeton University Press.
- Petersen, M. A. (2004). Information: Hard and Soft. Working Paper.
- Premack, D., Woodruff, G. (1978). Does the Chimpanzee Have a Theory of Mind?. *Behavioral and Brain Sciences*, 1 (4), 515–526.
- Ross, S. A. (1973). The Economic Theory of Agency: The Principal’s Problem. *American Economic Review*, 63 (2), 134–139.
- Schutz, A. (1932). *The Phenomenology of the Social World*. Northwestern University Press.
- Singer, T., Fehr, E. (2005). The Neuroeconomics of Mind Reading and Empathy. *American Economic Review*, 95 (2), 340–345.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., Moll, H. (2005). Understanding and Sharing Intentions: The Origins of Cultural Cognition. *Behavioral and Brain Sciences*, 28 (5), 675–735.
- Weber, M. (1979). *Economy and Society: An Outline of Interpretive Sociology*. University of California Press.

Acknowledgements

We thank (without implicating) Michael Donadelli, Michele Crescenzi, and Birger Wernerfelt for a number of helpful discussions. JL acknowledges support from the EPSRC, UK.

