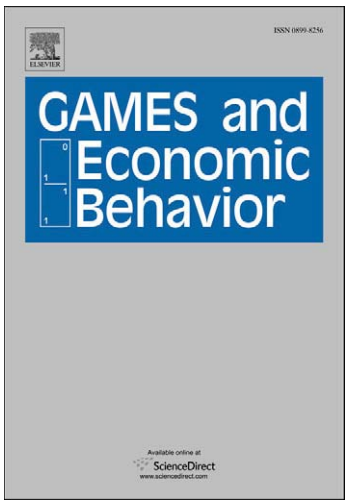

Overconfidence and Moral Hazard

Leonidas Enrique de la Rosa

PII: S0899-8256(11)00069-8
DOI: 10.1016/j.geb.2011.04.001
Reference: YGAME 1911

To appear in: *Games and Economic Behavior*

Received date: 12 June 2009

Please cite this article in press as: de la Rosa, L.E. Overconfidence and Moral Hazard. Games Econ. Behav. (2011), doi:10.1016/j.geb.2011.04.001

Games and Economic Behavior ●●●●, ●●●, ●●●

Overconfidence and Moral Hazard

Leonidas Enrique de la Rosa

Highlights

► We study the effects of overconfidence on incentive contracts with moral hazard. ► An overconfident agent tends to prefer higher-powered incentives. ► Lower-powered incentives are sufficient to induce effort. ► There are possible efficiency gains stemming from agent overconfidence. ► An increase in optimism or overconfidence increases the implemented effort level.

Overconfidence and Moral Hazard[☆]

Leonidas Enrique de la Rosa

School of Economics and Management, Aarhus University, Building 1322, DK–8000 Aarhus C, Denmark.

Abstract

In this paper, I study the effects of overconfidence on incentive contracts in a moral-hazard framework. Agent overconfidence can have conflicting effects on the equilibrium contract. On the one hand, an optimistic or overconfident agent disproportionately values success-contingent payments, and thus prefers higher-powered incentives. On the other hand, if the agent overestimates the extent to which his actions affect outcomes, lower-powered incentives are sufficient to induce any given effort level. If the agent is moderately overconfident, the latter effect dominates. Because the agent bears less risk in this case, there are efficiency gains stemming from his overconfidence. If the agent is significantly overconfident, the former effect dominates; the agent is then exposed to an excessive amount of risk, and any gains arise only from risk-sharing under disagreement. An increase in optimism or overconfidence increases the effort level implemented in equilibrium.

Key words: overconfidence, heterogeneous beliefs, moral hazard

JEL: A12, D82, D86

1. Introduction

The principal-agent model has been widely used to model incentives in organizations (for a survey, see Prendergast (1999)). It provides insight into the forces that shape incentive contracts, based on the assumption of full rationality. There is extensive psychological evidence that people are overconfident about their ability and future prospects.[1] This behavioral characteristic seems

[☆]An early version of this paper is part of my Ph.D. dissertation at the University of California, Berkeley. I am particularly grateful to Peter Ove Christensen, Benjamin Hermalin, Matthew Rabin, anonymous referees, and especially Botond Koszegi for their insightful comments. I am also very grateful to Robert Anderson, Ulrike Malmendier, Dayanand Manoli, John Morgan, Terrance Odean, Yuliy Sannikov, Chris Shannon, Adam Szeidl, and seminar participants at Aarhus University, UC Berkeley, CIDE, EPGE–FGV, Melbourne Business School, Oberlin College, and Queen's University for helpful comments and discussion.

Email address: delarosa@econ.au.dk (Leonidas Enrique de la Rosa)

[1]In this paper, I use the term "overconfidence" in reference to overestimating the probability of favorable outcomes, in particular those that depend on the agent's actions. Some authors refer to this type of self-serving bias as unrealistic optimism, positive self-image, overoptimism, or simply optimism, while others share my use of the term overconfidence. When discussing ability and the repercussions of one's actions, overconfidence is a more appropriate term than

particularly important in the context of an agency problem, since the agent's ability and stochastic distribution of outcomes conditional on the agent's actions are crucial in such a setting. The purpose of this paper is to study the effects of agent overconfidence in a moral-hazard model, particularly its effect on the equilibrium incentive contract.

The standard moral-hazard model describes the problem facing a principal who hires a risk-averse agent; consider for example the owner of a car dealership who hires a salesman. If the owner cannot monitor the effort that the salesman puts into a potential sale, which affects profit, she will offer him an incentive contract (e.g. consisting of salary and commission). The salesman will choose to exert effort because he knows that, by doing so, he will increase the probability of a sale and thus of earning a commission. The consequence of the moral-hazard problem is that the owner must trade off insurance and incentives: increasing the commission component gives incentives to the salesman to exert more effort, but a risk-averse salesman will have to be compensated for the inherent risk he faces. It is customary to assume that the parties hold identical beliefs regarding how the salesman's effort affects the probability distribution of a sale. Relaxing this assumption affects the incentive-insurance tradeoff in interesting ways.

We will introduce heterogeneous beliefs of which principal and agent are aware: they "agree to disagree." The owner believes the salesman is overconfident, while the salesman is convinced his expectations are realistic. There are two dimensions on which the asymmetry of beliefs turns out to be important in the model: an overconfident salesman can overestimate the probability of success for any given effort level, which we will refer to as *optimism*, and he can overestimate the marginal contribution of his effort to the probability of success, which we will refer to as *overconfidence*. Because of the parties' awareness of the asymmetry in beliefs, there are no signaling or screening concerns in this model, which allows us to isolate the effects of overconfidence on optimal contract design from its consequences in terms of adverse selection.

Agent overconfidence can have conflicting effects on the equilibrium contract. On the one hand, when the salesman is overconfident, a lower commission is sufficient to induce any effort level other than the cost-minimizing level of effort. I refer to this as the ***incentive effect*** of overconfidence. It pushes the equilibrium contract towards lower-powered incentives. On the other hand, because an overconfident salesman overestimates the probability of a sale, he finds

optimism, which we will reserve for overestimating the probability of outcomes independent of the agent's action. As we will see, these two types of bias have distinct effects on contracting. In the psychological literature, there are two other uses of the term overconfidence: the "better-than-average" effect (overestimating their ranking with respect to others) and overestimating the precision of one's beliefs. Moore and Healy (2008) discuss these three uses and propose a model in which the three types of overconfidence arise.

high-commission incentive contracts more attractive than a "realistic" agent would. Because the owner believes that she will pay the commission relatively infrequently, she finds such a contract—with a higher commission and a lower base salary—an inexpensive way of hiring the agent. This consequence of the divergence in evaluating payments is the ***wager effect*** of overconfidence. It pushes the equilibrium contract towards higher-powered incentives. Given that agent optimism does not affect the marginal distribution of outcomes conditional on effort, agent optimism has only a wager effect.

The degree of overall agent bias determines which of these effects dominates in the case that effort is implemented in equilibrium. The incentive effect dominates if the agent is slightly overconfident overall, so the principal provides more insurance without destroying incentives. If the agent is significantly overconfident overall, incentive provision becomes secondary to the fact that principal and agent value outcome-contingent payments differently. The wager effect dominates in this case, and greater agent overconfidence results in higher-powered incentives in equilibrium. Because of these potentially conflicting effects, the power of incentives of the equilibrium contract depends both on the degree of overall bias and the level of agent optimism and overconfidence. In contrast, the level of effort implemented by the equilibrium contract unambiguously increases with both optimism and overconfidence, since both the incentive and the wager effects make higher levels of effort less costly to implement.

Another interesting result is that agent overconfidence can be beneficial in terms of efficiency. An agent who is slightly overconfident receives more insurance than an agent who holds realistic beliefs. Providing more insurance to the risk-averse agent reduces the cost of agency, given that the principal is risk-neutral. If the principal receives all the gains from trade, the agent can only stand to lose from holding inaccurate beliefs. However, if the agent receives the gains from trade, the efficiency gains arising from a slight level of overconfidence might actually result in welfare gains.

There is a growing literature that takes the possibility of overconfidence into account. Studies of overconfidence in diverse settings include Bernardo and Welch (2001) in informational cascades, Gervais and Goldstein (2007) in a model of teams with complementarities, Goel and Thakor (2008) in tournaments, and Manove and Padilla (1999) in a screening model in which there are complementarities of effort among several agents. Within the agency literature, several studies focus on the effects of overconfidence in the presence of adverse selection; Maskin and Tirole (1990) and Maskin and Tirole (1992) introduce private information held by the principal regarding the extent to which she values the agency relationship in an adverse-selection model, de Meza and Southey (1996) develop a self-selection model which results in the most overconfident entrepreneurs posting the highest collateral, and Villeneuve (2000) and Koufopoulos (2008) look at adverse selection in

insurance markets with informed principals and informed agents, respectively. Gervais et al. (2009) model the possibility of overconfidence within a framework of information acquisition; the manager has access to a private signal and may overestimate the extent to which the signal is informative about true project profitability rather than noise.

This paper is most closely related to contemporaneous work on overconfidence regarding the effects of an agent's actions on outcome distribution in a moral-hazard setting. Santos-Pinto (2008) studies the implications of overconfidence in terms of profit for the principal when she makes a take-it-or-leave-it offer to the agent. He characterizes sufficient conditions under which increasing overconfidence will translate into higher profit but is not concerned with the shape of the incentive contract. Adrian and Westerfield (2009) specifically study the shape of the contract in a dynamic setting in which only the principal updates her beliefs. They find the possibility of divergence in the path of incentive contracts even though beliefs converge. As they point out, their model is one of heterogeneous beliefs about the project's profitability, rather than agent overconfidence about the effectiveness of his actions. Their results are, in the terms of this paper, regarding the wager effect of agent optimism in a dynamic setting.

The main contribution of this paper is to analyze the effects of agent overconfidence on the shape of incentive contracts. The setting we study allows for the explicit characterization of equilibrium incentive contracts in the presence of agent overconfidence, and thus for distinguishing two possibly conflicting effects of agent overconfidence in a moral-hazard setting. This distinction is important, since the wager and incentive effects have opposite effects in terms of the power of incentives of the equilibrium contract but reinforcing effects in terms of the implemented effort level.

The rest of the paper is organized as follows. Section 2 introduces the main assumptions and setup of the model, devoting special attention to the assumption that principal and agent *knowingly* hold asymmetric beliefs. Section 3 develops the main results of the model in a simple setting in which the agent's action choice is binary. Section 4 extends the analysis to a continuum of effort levels in the agent's choice set. Section 5 concludes.

2. Framework

The main assumption which makes this model differ from the standard moral-hazard model is that principal and agent hold heterogeneous beliefs regarding the distribution of outcomes, of which both are aware. Therefore, principal and agent do not update their beliefs along the gamepath (principal and agent simply "agree to disagree"). Because this assumption is crucial to the results of this paper, I will pause to discuss its validity.

There are both empirical and methodological reasons for assuming that parties do not fully

update their beliefs upon learning the beliefs held by others. This assumption is suitable for an agency framework, where only two economic agents interact. If there is disagreement about the agent's ability, the agent can convince himself that "I know myself better than anybody else," while the principal can discount the agent's beliefs since "everyone thinks they're better than average." This kind of arguments allow both agent and principal to rationalize not revising their beliefs. Consider, for example, the extreme situation in which the principal judges the agent's ability according to the population mean, knowing that agents tend to be overconfident. If she believes that the agent's beliefs are independent of his true ability[2], she will disregard those beliefs as uninformative. In this scenario, the principal's beliefs are independent of the individual agent's ability, so the agent can also disregard them as uninformative. Principal and agent have nothing new to say to each other in terms of the agent's true ability. Although this example is extreme, the important assumption I maintain is that some heterogeneity in beliefs persists after allowing the participants to update their information.

Heterogeneous posterior beliefs can result from differing prior beliefs. Morris (1995) discusses the assumption of heterogeneous priors in the context of economic models, and makes a case for allowing this possibility. Van den Steen (2004) models how overconfidence can arise from heterogeneous priors when individuals have a choice over projects. An alternative explanation involves errors in processing information (i.e. cognitive bias). If players update their beliefs in a non-Bayesian way, their posterior beliefs will differ even if all private information is revealed in equilibrium. Gervais and Odean (2001) show that if individuals overweigh success and underweigh failure when updating their beliefs about their own ability, they will "learn" to be overconfident. Bénabou and Tirole (2002) model a self-deception game in which multiple equilibria regarding the level of overconfidence may arise. Eyster and Rabin (2005) show that if players in a private-information game fail to interpret other players' actions as conveyors of private information, asymmetric posterior beliefs will survive even in fully-separating equilibria of the game.

In discussing the results, I focus mainly on the case of agent overconfidence overall: the agent holds overly optimistic beliefs, relative to the principal, regarding the probability of success of the project. The analysis, however, accommodates the possibility of a relatively underconfident or pessimistic agent, the implications of which are also discussed in the paper. This emphasis is consistent with research in the field of psychology: individuals tend to overestimate the probability of favorable events, and such bias is more pronounced when they have some control over the

[2](Cooper et al., 1988, p. 105), in an empirical study of entrepreneurs' preceptions of their chances of success, found that perceived odds of success were statistically unrelated to what they labeled as "objective predictors" of success.

likelihood of those events. Weinstein (1980) found that students were overly optimistic about the likelihood of good or bad events happening to them relative to same-gender students in their school—such as enjoying their post-graduation job or attempting suicide. He also found that the degree of such "unrealistic optimism" depended on, among other things, a notion of control over the likelihood of a given event. Taylor and Brown (1988) present a review of literature in psychology that supports the view that, in general, individuals' assessment of their own abilities, talents, and social skills are overly optimistic. They also note that overconfidence correlates with measures of mental health; it is mildly depressed individuals who hold realistic beliefs. Support for the case of overconfidence can also be found in the business and economics literature. Larwood and Whittaker (1977) found company managers to be unrealistically optimistic about the future performance of their firms relative to the competition. Cooper et al. (1988), based on a survey of nearly three thousand entrepreneurs, report that entrepreneurs are notably optimistic about their chances of success when setting up a business. Evidence from experimental economics supports the case for overconfidence as well: Camerer and Lovallo (1999), for example, find that there is excess entry into a hypothetical capacity-constrained market when participants' later payoffs depend on skill, but not when they depend on chance. These studies suggest that agents not only hold overconfident beliefs, but also act on them. The fact that the agent is directly in control of the actions that affect outcome distribution points to agent overconfidence (relative to the principal). (Van den Steen, 2004, p. 1144) uses the example of a company's CEO with decision control over strategies to make this point: in his model, "the CEO will be overoptimistic about the probability of success of his strategy, according to that shareholder," because it is the CEO who picked his strategy.

The other assumptions of the model are in line with the standard treatment of moral hazard. Assume there is a project that can be undertaken by a principal and an agent if they decide to enter a contractual relationship. There are two possible outcomes: the project can succeed or fail.[3] The project yields revenue x_0 if it fails, and revenue $x_1 > x_0$ if it succeeds. The probability of success of the project depends on a non-contractible action $e \in [0, 1]$ chosen by the agent, which can be interpreted as his choice among effort levels.

The principal's utility is expected revenue from the project net any payments made to the agent (the principal is risk neutral). The agent's utility is separable in money and effort, so that his utility can be written as

$$u(s) - c(e),$$

where $u(\cdot)$ denotes his money-utility of receiving payment s from the principal and $c(\cdot)$ denotes

[3]This simplifying assumption allows for an explicit characterization of the equilibrium incentive contract.

the disutility to the agent from exerting effort e. I assume that $u : \mathbb{R} \to \mathbb{R}$ has full range, and that it is continuous and twice continuously differentiable, with $u' > 0$ and $u'' < 0$ (the agent is risk averse). $c : [0, 1] \to \mathbb{R}$ will be characterized separately in the case of binary and continuous actions.

As previously noted, principal and agent knowingly hold asymmetric beliefs regarding the probability of success of the project. The principal believes that, conditional on the agent choosing effort level e, the project will succeed with probability $\Pr(x_1 \mid e) = q + ve$. Let a tilde denote the agent's beliefs: he believes that the conditional probability of success is $\widetilde{\Pr}(x_1 \mid e) = \tilde{q} + \tilde{v}e$. This particular parameterization will subsequently prove to be useful for the analysis, because it highlights the two dimensions (levels and differences) in which the asymmetry in beliefs is relevant in the model. Intuitively, allowing for a more general specification, it is only the marginal increase in the probability of success that affects the agent's choice of effort given any incentive scheme. Given an incentive scheme and the corresponding effort choice, however, it is the overall probability of success that determines the agent's expected utility of accepting a contract. The parameters q, $\tilde{q}$, v, and $\tilde{v}$ are assumed to be positive; the probability of success of the project is perceived by both parties to be increasing in effort. Beliefs are also restricted to $q + v < 1$ and $\tilde{q} + \tilde{v} < 1$.[4]

There are two ways in which the beliefs held by principal and agent can differ. The agent is said to be *optimistic* if $\tilde{q} > q$. The agent is said to be *overconfident* if $\tilde{v} > v$; he overestimates the marginal contribution of his effort to the probability of success relative to the principal's beliefs. The agent is said to be *overconfident overall* if $\widetilde{\Pr}(x_1 \mid e) > \Pr(x_1 \mid e)$ for all $e \in [0, 1]$, i.e., if $\tilde{q} > q$ and $\tilde{q} + \tilde{v} > q + v$.

The possibility of agent underconfidence ($\tilde{v} < v$) is consistent with overall overconfidence and may be relevant according to some views regarding self-enhancing biases. (Hoorens, 1993, pp. 131–132) notes that most self-enhancing biases seem to be motivated by a desire to see oneself as particularly "good" and a consequent perception of superiority. A sense of superiority might lead an agent to believe that the probability of success of a project in which he engages is very high independent of effort level (a very high $\tilde{q}$) and to thus underestimate the value of his effort ($\tilde{v} < v$).[5] The agent's beliefs about the value of effort affect his perception of the rewards to

[4]The assumption that $\tilde{q} + \tilde{v} < 1$ avoids the possibility of a trivial forcing contract—one that infinitely punishes the agent in case of project failure and thus trivially implements effort at first-best cost. Assuming $q + v < 1$ (so that principal and agent agree on the subset of outcomes that occur with probability zero) avoids the possibility that the principal can unboundedly increase the agent's perceived expected utility at no cost to herself, since success is a probability zero event.

[5]Imagine, as an extreme example, an agent who believes he has the "Midas touch": just because he is involved, the enterprise must succeed. This agent is overall very overconfident in the sense that he always overestimates the probability of success of the project, but at the same time he underestimates the relevance of his effort to increase

effort of a given incentive contract. Even though I am partial to interpret the evidence regarding overconfidence as pointing to overconfidence about the value of effort being the main force behind overall overconfidence, it is important to note this possibility.

The solution concept used is subgame-perfect Nash equilibrium: at every decision node of the game, the relevant player chooses an optimal response, even if she had expected not to reach that node in equilibrium. Without loss of generality, we restrict our attention to contract offers of the form $\langle s_1, s_0 \rangle$—a schedule of outcome-contingent payments to the agent—given that project outcome is the only mutually-observable signal in the model.[6]

3. Binary Action Space

Consider the standard agency framework under which one principal can make a take-it-or-leave-it contract offer to one agent. Assume the agent has two actions to choose between; $e \in \{0, 1\}$. This assumption will allow us to study the incentive and wager effects independently. In Section 4, we allow for a set of (continuous) actions available to the agent to show that this simplification is not driving the main results. A straightforward way to interpret this two-action space is that the agent can simply choose whether or not to exert effort. We normalize the cost of not exerting effort to zero, so that $c(0) = 0$ and $c(1) = c$.

The timing of the model is depicted in Figure 1. First, the principal makes a take-it-or-leave-it contract offer to the agent. The agent can accept or reject the offer. If he accepts it, he chooses whichever action maximizes his perceived expected utility given the terms of the contract. The outcome of the project is realized and observed by both parties. Payoffs are then distributed according to the provisions in the contract, and the agency relationship ends. If the agent rejects the principal's offer, both will receive utility according to their outside option. We will assume that the agent's outside option is exogenous and independent of his overconfidence.[7] The principal's

the probability of success.

[6] See Holmström (1979) for a discussion about observability and contracting under moral hazard. Because of the agent's risk aversion, it is in general not optimal to introduce unnecessary "noise" to the payment structure. If signals besides project outcome are observable by both parties, the terms of the equilibrium contract may be contingent on those as well. Under asymmetric beliefs, if there is some signal that the agent believes to be correlated with his effort, the principal can reduce the cost of implementing effort by offering payments that are also contingent on this signal, even if she believes it to be completely uninformative. This pure side-betting has little insight to offer in understanding the results we discuss.

[7] This assumption allows us to isolate the effect that overconfidence, restricted to the probability of success of the project, has on the equilibrium contract. If the agent were also overconfident about his outside option, he would demand a higher perceived expected utility in order to accept any given offer by the principal.

outside option does not affect the equilibrium contract as long as the agency relationship yields sufficient surplus for her to engage in it; we assume this to be the case. When evaluating prospects, principal and agent evaluate any given contract according to their own beliefs.

The principal wishes to maximize her expected profit which, given contract $\langle s_1, s_0 \rangle$ accepted by the agent in equilibrium and conditional on each of the agent's possible effort levels, is:

$$\mathbb{E}\left[\pi \mid e\right] = (q + ve)\left(x_1 - s_1\right) + \left[1 - (q + ve)\right]\left(x_0 - s_0\right).$$

Let $\underline{u}$ denote the agent's perceived expected utility from his outside option. The principal's contract offer, if it is to be accepted by the agent, must provide him with perceived expected utility no lower than $\underline{u}$. This participation, or "individual-rationality"—IR—constraint restricts the possible optimal contract offers to those that satisfy

$$(\tilde{q} + \tilde{v}e)\, u\left(s_1\right) + \left[1 - (\tilde{q} + \tilde{v}e)\right] u\left(s_0\right) - ce \geq \underline{u}, \tag{IR}$$

where e is the action (freely chosen by the agent) that the principal wishes to implement.

Finally, after accepting the principal's offer, the agent's objective is to maximize his expected utility when choosing how much effort to exert; recall that effort is non-contractible. Given contract $\langle s_1, s_0 \rangle$, the agent's expected utility conditional on his choice of effort is:

$$\tilde{\mathbb{E}}\left[u\left(s_x\right) \mid e\right] - ce = (\tilde{q} + \tilde{v}e)\, u\left(s_1\right) + \left[1 - (\tilde{q} + \tilde{v}e)\right] u\left(s_0\right) - ce.$$

We can now turn to characterizing the equilibrium contract. We will follow the method proposed by Grossman and Hart (1983) to compute the second-best optimal incentive scheme. We calculate the second-best contract that implements each possible action (in our case effort or no effort), which would allow the principal to calculate the second-best cost of implementing each action. The principal would then simply choose to implement the action which yields a higher expected profit.

3.1. First-Best Implementation

As a preliminary step, it is useful to characterize the first-best implementation of any given effort level. One can think of first-best implementation in terms of verifiable actions: the principal can specify an effort level she wishes to implement together with a set of payments. The principal need only concern herself with the agent's participation constraint and provide a "punishment" payment if the agent exerts any level of effort other than the one she wishes to implement. In the case of homogeneous beliefs, first-best implementation with a risk-averse agent and a risk-neutral principal is a contract that fully insures the agent and satisfies his participation constraint, i.e., the agent is paid the same amount whether the project succeeds or fails, and this amount covers his

opportunity cost and his disutility cost of effort (if any). Allowing for heterogeneous beliefs, optimal risk-sharing requires that the agent be exposed to risk in equilibrium—he will hold a "long" position in the project if his beliefs regarding success are optimistic relative to the principal's beliefs, and a "short" position in the project if he is pessimistic relative to the principal.

Proposition 1. *The first-best contract that implements effort level* $e \in \{0,1\}$, $\langle s_1^{FBe}, s_0^{FBe} \rangle$, *is characterized by the conditions*

$$\frac{\tilde{q}+\tilde{v}e}{1-(\tilde{q}+\tilde{v}e)}\frac{u'\left(s_1^{FBe}\right)}{u'\left(s_0^{FBe}\right)} = \frac{q+ve}{1-(q+ve)}$$

and $(\tilde{q}+\tilde{v}e)\,u\left(s_1^{FBe}\right)+[1-(\tilde{q}+\tilde{v}e)]\,u\left(s_0^{FBe}\right)-ce=\underline{u}$.

All proofs are relegated to the appendix. Note that in the case of homogeneous beliefs, $\frac{\tilde{q}+\tilde{v}e}{1-(\tilde{q}+\tilde{v}e)} = \frac{q+ve}{1-(q+ve)}$, which gives us the result that under homogeneous beliefs $s_1^{FBe} = s_0^{FBe} \equiv s^{FBe}$ such that $u\left(s^{FBe}\right) = \underline{u} + ce$. Proposition 1 is simply an extension of the Borch (1962) rule for Pareto-optimal risk-sharing allowing for heterogeneous beliefs. Intuitively, a relatively optimistic agent is willing to wager on success against the (relatively pessimistic) principal. We can use the first-best contract under homogeneous beliefs, with $s_1^{FBe} = s_0^{FBe}$, as a benchmark to understand the intuition behind optimal-risk sharing under heterogeneous beliefs. Because at that point the marginal cost of bearing additional risk is zero for the agent, principal and agent evaluate marginal changes in payments based on their effect only in terms of expected payment.[8] Consider, then, an increase in the success-contingent payment, coupled with a decrease in the failure-contingent payment, that leaves expected payment unchanged according to the agent's beliefs. The principal expects to pay the success-contingent payment less often than the agent expects to receive it, so *according to her beliefs* such deviation yields a lower expected payment to the agent, and thus higher expected profit. Evaluating at the riskless contract, there is a first-order gain for the principal from a lower expected payment and only a second-order loss from having to compensate the agent for higher risk exposure. Therefore, a relatively optimistic agent will bear risk under the first-best contract. For analogous reasons, so will a relatively pessimistic agent: he will receive higher payment contingent on project failure. Figure 2 shows the first-best contract that implements no effort given homogeneous beliefs (A) and given agent optimism (B).[9]

[8] Formally, $d\tilde{\mathbb{E}}\left[u\left(s_x\right)\right] = (\tilde{q}+\tilde{v}e)\,u'\left(s_1\right)ds_1 + [1-(\tilde{q}+\tilde{v}e)]\,u'\left(s_0\right)ds_0$. Starting from a riskless contract with $s_1 = s_0$, a marginal change in s_1 and s_0 that leaves expected payment constant according to the agent's beliefs—one such that $(\tilde{q}+\tilde{v}e)\,ds_1 + [1-(\tilde{q}+\tilde{v}e)]\,ds_0 = 0$—does not affect expected utility. At a riskless contract, given $s_1 = s_0$, $\frac{d\tilde{\mathbb{E}}[u(s_x)]/ds_1}{d\tilde{\mathbb{E}}[u(s_x)]/ds_0} = \frac{\tilde{q}+\tilde{v}e}{1-(\tilde{q}+\tilde{v}e)}$ and $\frac{d\mathbb{E}[\pi]/ds_1}{d\mathbb{E}[\pi]/ds_0} = \frac{q+ve}{1-(q+ve)}$.

[9] The graphical analysis gained much from suggestions made by two anonymous referees.

Corollary 1. *If agent preferences are such that* $\frac{1}{u'(s)}\left(-\frac{u''(s)}{u'(s)}\right)$ *is a non-decreasing function, the first-best contract,* $\langle s_1^{FBe}, s_0^{FBe}\rangle$*, exhibits power of incentives that increases in the degree of agent optimism and overconfidence:* $\frac{d}{d\tilde{q}}\left(u\left(s_1^{FBe}\right)-u\left(s_0^{FBe}\right)\right)>0$ *for* $e\in\{0,1\}$ *and* $\frac{d}{d\tilde{v}}\left(u\left(s_1^{FBe}\right)-u\left(s_0^{FBe}\right)\right)>0$ *for* $e=1$.

Because of the disagreement between principal and agent regarding the probability of success of the project, an optimistic agent is exposed to more risk than a "realistic" agent would be exposed to. This ***wager effect*** will tend to push the equilibrium contract towards higher-powered incentives in the case of agent optimism (larger negative power of incentives in the case of agent pessimism).[10] The wager effect is illustrated in Figure 3. As noted, absent moral-hazard concerns the equilibrium contract allows for *ex-ante* Pareto-optimal risk sharing. In the identical-beliefs case, this implies that the risk-neutral principal will absorb all of the risk. In the heterogeneous-beliefs case, it implies that the agent bears risk in proportion to the disagreement in beliefs.

The exception hinted at in Corollary 1 by the condition that $\frac{1}{u'(s)}\left(-\frac{u''(s)}{u'(s)}\right)$ be a non-decreasing function is worth clarifying. $\frac{1}{u'(s)}$ is increasing in s, but assuming the sufficient condition that $\left(-\frac{u''(s)}{u'(s)}\right)$ be a non-decreasing function means assuming non-decreasing absolute risk aversion, which seems counterintuitive. Consider the comparative statics in the case of a very optimistic agent. An increase in agent optimism immediately slackens the agent's participation (IR) constraint, given the increase in the likelihood of the event under which he is receiving a higher payment. This opens the possibility (depending on the agent's preferences) for the principal to reduce both payments, s_1 and s_0, while still satisfying the (IR) constraint. If the agent's absolute risk aversion increases sharply as a result of the reduction in wealth, the power of incentives of the second-best contract might indeed fall. For reference, note that the family of HARA utility functions (exponential, logarithmic, and power utility functions) do satisfy the condition that $\frac{1}{u'(s)}\left(-\frac{u''(s)}{u'(s)}\right)$ be a non-decreasing function while exhibiting non-increasing absolute risk aversion.

3.2. Second-Best Implementation

We now turn to second-best implementation. The contract must be incentive compatible: the contract terms must be such that the agent finds it in his best interest to exert the level of effort that the principal wishes to implement. We will assume that whenever indifferent, the agent exerts the level of effort chosen by the principal.

Consider second-best implementation of effort, which is the case of main interest in the standard model with homogeneous beliefs. Incentive compatibility in this case requires

[10] We will refer to the differential $u(s_1)-u(s_0)$ as the contract's *power of incentives*.

$$(\tilde{q}+\tilde{v})\,u(s_1)+[1-(\tilde{q}+\tilde{v})]\,u(s_0)-c\geq \tilde{q}u(s_1)+[1-\tilde{q}]\,u(s_0).$$

We can rewrite the incentive-compatibility constraint above as

$$\tilde{v}\,(u(s_1)-u(s_0))\geq c. \tag{IC}$$

Intuitively, the perceived expected utility gain for the agent from exerting effort (receiving excess utility $u(s_1)-u(s_0)$ with additional probability $\tilde{v}$), must be no less than his disutility from exerting effort (c). Note that the power of incentives necessary to induce effort is decreasing in $\tilde{v}$. It is, however, independent of $\tilde{q}$: only the agent's overconfidence, not his degree of optimism, affects his perception of the rewards from effort exertion given some incentive scheme.

If principal and agent hold identical beliefs, the second-best contract will be characterized by the binding individual-rationality (IR) and incentive-compatibility constraints (IC). There is a tradeoff between incentives and insurance: absent the incentive-provision problem, the principal would offer full insurance to the agent (i.e. reduce the power of incentives). Implementing effort requires that the agent be exposed to a discrete amount of risk: enough so that the increase in terms of expected utility from a higher probability of receiving the success-contingent payment compensates the agent for the disutility of exerting effort. The efficiency loss that arises, given the agent's risk aversion, is referred to as the cost of agency; if the agency relationship was not necessary, this cost would be avoided (e.g., if the risk-neutral principal could undertake the project on her own and carry out the agent's task).

Given this incentive-insurance tradeoff, the contract analogous to the identical-beliefs second-best contract that implements effort will be a useful reference when we allow for heterogeneous beliefs.

Definition 1. *Let $\langle \bar{s}_1, \bar{s}_0\rangle$ denote the contract that satisfies (IR) and (IC) with equality:*

$$(\tilde{q}+\tilde{v})\,u(\bar{s}_1)+[1-(\tilde{q}+\tilde{v})]\,u(\bar{s}_0)-c=\underline{u},$$
$$\tilde{v}\,(u(\bar{s}_1)-u(\bar{s}_0))=c.$$

This contract will in fact be the second-best contract that implements effort when the beliefs held by the agent differ only slightly from the principals' beliefs. It is the level of overall bias (taking into account both optimism and overconfidence) that will determine whether or not the $\langle \bar{s}_1, \bar{s}_0\rangle$ contract is second-best.

Definition 2. *The agent is said to be* slightly overconfident overall *if*

$$\frac{\tilde{q}+\tilde{v}}{1-(\tilde{q}+\tilde{v})}\frac{u'(\bar{s}_1)}{u'(\bar{s}_0)}\leq \frac{q+v}{1-(q+v)}.$$

Conversely, the agent is said to be significantly overconfident overall *if*

$$\frac{\tilde{q}+\tilde{v}}{1-(\tilde{q}+\tilde{v})}\frac{u'(\bar{s}_1)}{u'(\bar{s}_0)} > \frac{q+v}{1-(q+v)}.$$

The particulars about this definition become apparent in the proof of Proposition 2 below.[11] The intuition follows from the question of whether optimal risk sharing under disagreement requires higher- or lower-powered incentives than effort implementation requires. If the agent is only slightly overconfident overall, optimal risk sharing mandates for lower-powered incentives (i.e. more insurance for the agent) than required for effort implementation. The intuition from the identical-beliefs setting carries over in this case: a principal cannot provide the level of insurance required by optimal risk sharing without destroying the incentives for the agent to exert effort. Note that, given $\bar{s}_1 > \bar{s}_0$, the condition of slight overconfidence overall holds strictly when $\tilde{q}+\tilde{v} = q+v$, and thus the condition also holds for some values $\tilde{q}+\tilde{v} > q+v$ (i.e. there is a non-empty range of positive overall bias under the definition of slight overconfidence overall). Any degree of negative overall bias is subsumed within the definition of slight overconfidence overall.

Proposition 2. *If the agent is slightly overconfident overall, the second-best contract that implements effort is* $\langle \bar{s}_1, \bar{s}_0 \rangle$.

If the agent is only slightly overconfident overall, providing more insurance to the agent (by decreasing s_1 and increasing s_0) would destroy the incentives for the agent to exert effort. Compensating the agent for bearing more risk would be too costly for the principal, so the second-best contract has power of incentives just high enough to implement effort. If the agent is slightly overconfident overall, the power of incentives required to implement effort is larger than what would be Pareto-optimal risk sharing under heterogeneous beliefs. Figure 4 illustrates this point. Providing sufficient incentives to exert effort is what drives the power of incentives in this case (i.e. the relevant (IC) constraint binds).

Corollary 2. *If the agent is slightly overconfident overall, the second-best contract that implements effort,* $\langle \bar{s}_1, \bar{s}_0 \rangle$*, exhibits power of incentives that decreases in the degree of agent overconfidence:* $\frac{d}{d\tilde{v}}\left(u(\bar{s}_1) - u(\bar{s}_0)\right) < 0$. *Agent optimism does not affect the power of incentives:* $\frac{d}{d\tilde{q}}\left(u(\bar{s}_1) - u(\bar{s}_0)\right) = 0$.

[11] The agent is said to be significantly overconfident overall if a marginal change in s_1 and s_0 from $\langle \bar{s}_1, \bar{s}_0 \rangle$ towards more risk for the agent (increasing s_1 and decreasing s_0), that at the same time leaves the agent's expected utility unchanged, increases the principal's expected profit.

Intuitively, when the agent is only slightly overconfident overall, his overconfidence ($\tilde{v} > v$) allows the principal to provide him more insurance without destroying incentives. This is the ***incentive effect*** of overconfidence, illustrated in Figure 5.

Any contract that implements effort exposes the agent to a discrete amount of risk, so as to give him sufficient incentives to exert effort. Even though an overconfident agent is willing to wager to some extent with the principal, the amount of risk he bears according to optimal risk-sharing is continuous in the degree of disagreement in beliefs (recall Figure 3). If the agent is only slightly overconfident overall, the amount of risk required by incentive provision is greater than the amount of risk he would optimally bear given disagreement. When the agent is only slightly overconfident overall, the incentive-insurance tradeoff present in the case of identical beliefs remains. The incentive effect therefore dominates the wager effect when the agent is slightly overconfident overall, and the power of incentives of the second-best contract depends solely on the agent's beliefs about the marginal contribution of effort to the probability of success ($\tilde{v}$). If the agent is overconfident, the principal can provide more insurance than if the agent were rational, thus at a lower cost of agency. Note that agent underconfidence will require higher-powered incentives to implement effort; an underconfident agent underestimates the expected compensation differential from exerting effort, and thus agent underconfidence will increase the cost of agency. Finally, the agent's optimism ($\tilde{q}$) affects the incentive contract only through the (IR) constraint; more agent optimism means it is cheaper in expected terms for the principal to satisfy the agent's participation constraint, but it does not affect the power of incentives. The principal's beliefs do not affect the contract at all in this case.

The degree of disagreement may be so large that the agent is significantly overconfident overall. If so, the second-best contract that implements effort exhibits excessively powerful incentives in the sense that they are more powerful than necessary to implement effort (i.e. the (IC) constraint no longer binds). Because of the wager effect, an agent who is significantly overconfident overall overestimates the probability of success to such an extent that he prefers a contract that rewards him handsomely for success and punishes him harshly for failure over the $\langle \bar{s}_1, \bar{s}_0 \rangle$ contract (which provides as much insurance as possible while implementing effort). In this case the first-best contract implementation of effort satisfies incentive compatibility.

Definition 3. *Let $\langle s_1^*, s_0^* \rangle$ denote the contract defined by first-best implementation of effort as in Proposition 1, i.e., the contract characterized by the conditions*

$$\frac{\tilde{q}+\tilde{v}}{1-(\tilde{q}+\tilde{v})}\frac{u'(s_1^*)}{u'(s_0^*)} = \frac{q+v}{1-(q+v)}$$

and $(\tilde{q}+\tilde{v})\,u(s_1^*) + [1-(\tilde{q}+\tilde{v})]\,u(s_0^*) - e = \underline{u}$.

Proposition 3. *If the agent is significantly overconfident overall, the first-best contract that implements effort, $\langle s_1^*, s_0^* \rangle$, is incentive compatible and thus also the second-best contract that implements effort.*

If $\tilde{q}+\tilde{v} \gg q+v$ (such that he is significantly overconfident overall), the agent is actually content to bear more risk than he would under the contract $\langle \bar{s}_1, \bar{s}_0 \rangle$. Because of the wager effect, if the agent is significantly overconfident overall, he judges the $\langle s_1^*, s_0^* \rangle$ contract to yield a significantly higher expected payment than the $\langle \bar{s}_1, \bar{s}_0 \rangle$ contract—so much higher that it compensates him for the excessive amount of risk he bears. This means that it is less expensive for the principal to implement effort by offering $\langle s_1^*, s_0^* \rangle$ rather than $\langle \bar{s}_1, \bar{s}_0 \rangle$. Given that effort is implemented by the first-best contract, there is no cost of agency in this case.

When the agent is significantly overconfident overall, the agent's bias in evaluating payments overshadows the incentive-insurance tradeoff present in the standard moral-hazard setting with homogeneous beliefs. Incentive provision becomes secondary to the wagering motive that arises from the difference in how principal and agent evaluate outcome-contingent payments. Figure 6 illustrates this point: the second-best contract that implements effort given significant agent overconfidence (B) has higher power of incentives than necessary to implement effort (i.e. the relevant (IC) constraint is slack).

Corollary 3. *If the agent is significantly overconfident overall, and agent preferences are such that $\frac{1}{u'(s)}\left(-\frac{u''(s)}{u'(s)}\right)$ is a non-decreasing function, the second-best contract that implements effort, $\langle s_1^*, s_0^* \rangle$, exhibits power of incentives that increases in the degree of agent optimism and overconfidence: $\frac{d}{d\tilde{q}}\left(u\left(s_1^*\right)-u\left(s_0^*\right)\right)>0$ and $\frac{d}{d\tilde{v}}\left(u\left(s_1^*\right)-u\left(s_0^*\right)\right)>0$.*

Figure 7 illustrates the second-best contract that implements effort (in boldface) as a function of the agent's beliefs about the value of his effort, $\tilde{v}$. The implications of Propositions 2 and 3 are evident: the incentive effect dominates when the agent is only slightly overconfident overall, and the wager effect dominates when the agent is significantly overconfident overall. As a consequence, the power of incentives of the second-best contract that implements effort is not a monotonic function of agent overconfidence.

Second-best implementation of no effort is the counterpart to second-best implementation of effort. Under heterogeneous beliefs, first-best implementation of no effort is incentive compatible and thus also second best, but only if the degree of agent optimism is not very large. Significant overconfidence overall allows second-best implementation of effort by the first-best contract, whereas significant optimism implies that the first-best contract that implements no effort exhibits such large

power of incentives that exerting no effort is not incentive-compatible. In other words, the wager effect implies such a difference in the success- and failure-conditional payments that the first-best contract would implement effort in a moral-hazard framework. Given that the intuition is the counterpart to second-best implementation of effort, we define the concepts and state the results without further discussion.

Definition 4. *Let $\langle s_{1*}, s_{0*}\rangle$ denote the contract defined by first-best implementation of no effort as in Proposition 1, i.e., the contract characterized by the conditions*

$$\frac{\tilde{q}}{1-\tilde{q}}\frac{u'(s_{1*})}{u'(s_{0*})} = \frac{q}{1-q}$$

and $\tilde{q}u(s_{1*}) + [1-\tilde{q}]\,u(s_{0*}) = \underline{u}$.

Definition 5. *Let $\langle \underline{s}_1, \underline{s}_0\rangle$ denote the contract that implements no effort and that satisfies the respective agent's participation and incentive-compatibility constraints with equality:*

$$\tilde{q}u(\underline{s}_1) + [1-\tilde{q}]\,u(\underline{s}_0) = \underline{u},$$
$$\tilde{v}\,(u(\underline{s}_1) - u(\underline{s}_0)) = c.$$

Definition 6. *The agent is said to be* slightly optimistic *if*

$$\frac{\tilde{q}}{1-\tilde{q}}\frac{u'(\underline{s}_1)}{u'(\underline{s}_0)} \leq \frac{q}{1-q}.$$

Conversely, the agent is said to be significantly optimistic *if*

$$\frac{\tilde{q}}{1-\tilde{q}}\frac{u'(\underline{s}_1)}{u'(\underline{s}_0)} > \frac{q}{1-q}.$$

Proposition 4. *If the agent is slightly optimistic, the first-best contract that implements no effort, $\langle s_{1*}, s_{0*}\rangle$, is incentive compatible and thus also the second-best contract that implements effort. If the agent is significantly optimistic, the second-best contract that implements no effort is $\langle \underline{s}_1, \underline{s}_0\rangle$.*

Corollary 4. *If the agent is slightly optimistic, and agent preferences are such that $\frac{1}{u'(s)}\left(-\frac{u''(s)}{u'(s)}\right)$ is a non-decreasing function, the second-best contract that implements no effort, $\langle s_{1*}, s_{0*}\rangle$, exhibits power of incentives that increases in the degree of agent optimism: $\frac{d}{d\tilde{q}}(u(s_{1*}) - u(s_{0*})) > 0$. Agent overconfidence does not affect the power of incentives:$\frac{d}{d\tilde{v}}(u(s_{1*}) - u(s_{0*})) = 0$.*

Corollary 5. *If the agent is significantly optimistic, the second-best contract that implements no effort, $\langle \underline{s}_1, \underline{s}_0\rangle$, exhibits power of incentives that decreases in the degree of agent overconfidence: $\frac{d}{d\tilde{v}}(u(\underline{s}_1) - u(\underline{s}_0)) < 0$. Agent optimism does not affect the power of incentives: $\frac{d}{d\tilde{q}}(u(\underline{s}_1) - u(\underline{s}_0)) = 0$.*

The second-best contracts provide the principal with the best way to implement effort or no effort. The principal need only compare her expected profit when implementing effort with her expected profit when implementing no effort to decide which is the optimal contract offer. Although this comparison is a simple mechanical process, we cannot find a closed-form solution for the effort level that is implemented in equilibrium without particular assumptions about the functional form of the agent's utility. We can, however, study the effects of changes in underlying parameters on the implemented effort level.[12] Of particular interest is the effect of changes in the agent's optimism and overconfidence. Because the incentive effect reduces the cost of implementing effort and the wager effect makes high-power-of-incentives contracts more attractive to the agent, higher levels of either optimism or overconfidence imply a larger set of parameters under which effort will be implemented in equilibrium.

Proposition 5. *Holding other parameters constant, if effort is implemented given agent's beliefs* $(\tilde{q}, \tilde{v})$*, then effort will be implemented when his beliefs are* $(\tilde{q}, \tilde{v}')$ *for any* $\tilde{v}' \geq \tilde{v}$ *or* $(\tilde{q}', \tilde{v})$ *for any* $\tilde{q}' \geq \tilde{q}$*.*

A higher level of overconfidence slackens the agent's participation constraint only if effort is implemented, which makes the contract that implements effort more attractive to the principal. When the incentive effect dominates, this effect is strengthened by the fact that higher overconfidence reduces the cost of agency. A higher level of optimism, on the other hand, slackens the agent's participation constraint whether effort is implemented or not. Because the agent is more invested in success under a contract that implements effort (because of the incentive-compatibility constraint), the gains from a slackening of the participation constraint are larger under the second-best contract that implements effort relative to the second-best contract that implements no effort.

Note that Proposition 5 does not imply effort is more likely to be implemented, in general, when dealing with an overall more overconfident agent, given the possibility of agent underconfidence coupled with overall overconfidence that we discussed at the end of Section 2.

[12] It is easy to see that in our framework effort implementation is less often optimal if $(x_1 - x_0)$ is reduced or if c is increased. More detailed analysis would require assuming some sort of functional form of agent utility. (Laffont and Martimort, 2002, pp. 160-162) discuss the optimality of second-best effort implementation in general and the comparative statics of whether effort is implemented or not by assuming a quadratic form for the inverse of the agent's utility function. In short, they find that second-best effort implementation is less often optimal if the agent's risk aversion, the disutility cost of effort, or the principal's inference problem are increased.

3.3. Welfare Analysis

Disagreement between principal and agent affects the implemented contract in an agency relationship. The equilibrium contract depends only on their beliefs, independent of whether or not either holds the "correct" beliefs. Evaluating welfare based solely on subjective beliefs, we can say that there are *ex-ante* Pareto gains from disagreement; recall that the (IR) constraint is always binding in equilibrium. This means that the agent's expected utility is equal to his outside option, and we can show that the principal's expected profit increases in the agent's level of optimism and overconfidence if effort is implemented (cf. Proposition 2 in Santos-Pinto (2008)). If no effort is implemented, the principal's expected profit increases in the degree of disagreement (allowing for both optimism and pessimism), while it decreases in agent overconfidence if the agent is significantly optimistic.

Proposition 6. *If effort is implemented, the principal's expected profit increases in both the agent's level of optimism and overconfidence. If no effort is implemented, the principal's expected profit increases in the agent's level of optimism or pessimism, for an optimistic or a pessimistic agent, respectively, and it decreases in agent overconfidence if the agent is significantly optimistic.*

Optimal "side-betting" when no effort is implemented explains that expected profit for the principal increases in optimism or pessimism; the agent receives a higher payment conditional on the event (success or failure) that he believes is more likely relative to the principal's beliefs. A larger degree of disagreement slackens the (IR) constraint since there is an increased probability of receiving the higher payment. Similarly, implementing effort requires a higher payment conditional on success. So increases in both agent optimism and overconfidence slacken the (IR) constraint, allowing for higher expected profit for the principal. Finally, if no effort is implemented and the incentive-compatibility constraint binds, the principal's expected profit decreases in agent overconfidence, since the incentive effect decreases the power of incentives away from Pareto-optimal risk sharing (side-betting).

There is another source of increased expected profit for the principal when effort is implemented: if the agent's beliefs are such that he is only slightly overconfident overall, the (IC) constraint is binding, and a higher level of agent overconfidence (but not higher optimism) slackens this constraint as well. This results in a reduction of the cost of agency, since the principal can provide more insurance to the risk-averse agent, which translates into higher expected profit for the principal.

Imagine that the principal holds the "correct" beliefs; we might want to ask how well the overconfident agent is faring. If we were to evaluate the agent's expected utility using the principal's beliefs, holding incorrect beliefs would always be detrimental to the agent: he would always receive

a lower "actual" expected utility than his outside option. If, however, the agent were to receive the gains from trade (as would be the case if two or more principals competed in making contract offers) he would in fact benefit from holding incorrect overconfident beliefs within the slight overconfidence overall range, since he would benefit from the reduction in the cost of agency.[13] Note that these potential gains arise from the fundamental tradeoff between insurance and incentives rather than the presence of production externalities (cf. Gervais and Goldstein (2007)).

4. Continuous Action Space

The purpose of this section is to show that, maintaining the structure of how optimism and overconfidence affect the probability of success, the results put forward in Section 3 are robust in the sense that they extend to a continuous action space.

Assume that the agent can choose some $e \in [0,1]$ if he accepts the principal's contract offer. We maintain the rather simple specification of the probability of success being a linear function of the level of effort, which did not sacrifice generality in the binary action case but is restrictive in the case of continuous actions, for tractability. The disutility cost of effort is a function $c : [0,1] \to \mathbb{R}^+$ which is assumed to be continuous and twice differentiable, with $c'(\cdot) > 0$ and $c''(\cdot) > 0$. This implies that the agent's effort level choice will be proportionately related to the contract's power of incentives as long as his choice is an interior solution to his perceived expected utility maximization problem. Assume that $c'(0) = 0$ and $\lim_{e\to 1} c'(e) = \infty$ so that it is, in fact, an interior solution whenever $s_1 > s_0$.[14]

The agent's individual-rationality constraint changes slightly to allow for a continuous cost-of-effort function;

$$(\tilde{q} + \tilde{\nu}e)\, u(s_1) + [1 - (\tilde{q} + \tilde{\nu}e)]\, u(s_0) - c(e) \geq \underline{u}, \qquad \text{(IR')}$$

where e denotes the action that the principal wishes to implement. If the agent accepts a given contract offer $\langle s_1, s_0 \rangle$, he will subsequently choose his effort level so as to maximize his perceived expected utility:

$$\max_{e\in[0,1]} (\tilde{q} + \tilde{\nu}e)\, u(s_1) + [1 - (\tilde{q} + \tilde{\nu}e)]\, u(s_0) - c(e).$$

Two of our assumptions allow us to apply the first-order approach in this setting: effort affects the probability of success linearly and $c''(e) > 0$ for all e imply that the agent's optimization problem

[13] This setting is analyzed in the working paper version of this article, de la Rosa (2007).

[14] Note that if the agent chooses a corner solution ($e = 0$ or $e = 1$), as long as his choice of effort remains at a given corner or shifts discretely to the other, the analysis reduces to a binary-action model as studied.

is strictly concave. The first-order condition for the agent's problem is

$$\tilde{v}\left[u\left(s_1\right)-u\left(s_0\right)\right]=c'\left(e\right), \tag{IC'}$$

which defines the agent's choice of effort after accepting contract offer $\langle s_1, s_0\rangle$. The incentive effect is apparent from this condition: a lower-powered incentive contract is sufficient to implement any given effort level e when the agent is overconfident about the value of effort.

As before, the agent's participation constraint (IR') must be binding in equilibrium. If it did not, the principal could marginally reduce both payments s_1 and s_0 while keeping the power of incentives $[u\left(s_1\right)-u\left(s_0\right)]$ constant (so as to implement the same effort level), increasing her expected profit and thus contradicting optimality.

In order to characterize the relationship between overconfidence, the power of incentives, and the implemented level of effort under the optimal contract, it is useful to reinterpret the principal's problem. Note that the binding participation constraint (IR') and the incentive-compatibility constraint (IC') characterize the second-best contract for the principal to implement any given effort level e. Given the agent's effort choice problem, and that the principal will optimally set the participation constraint to bind, the best contract that implements effort level e, $\langle s_1\left(e\right), s_0\left(e\right)\rangle$, is therefore implicitly defined by

$$u\left(s_1\left(e\right)\right)=\underline{u}+c\left(e\right)+\left[1-\left(\tilde{q}+\tilde{v}e\right)\right]\frac{c'\left(e\right)}{\tilde{v}} \text{ and}$$
$$u\left(s_0\left(e\right)\right)=\underline{u}+c\left(e\right)-\left(\tilde{q}+\tilde{v}e\right)\frac{c'\left(e\right)}{\tilde{v}}.$$

Taking this into account, we can reduce the principal's problem to

$$\max_{e\in[0,1]}\left(q+ve\right)\left(x_1-s_1\left(e\right)\right)+\left[1-\left(q+ve\right)\right]\left(x_0-s_0\left(e\right)\right),$$

where $s_1\left(e\right)$ and $s_0\left(e\right)$ are implicitly defined in the preceding expressions. We can see that the power of incentives of the second-best contract depends on the agent's beliefs, the agent's marginal cost of effort, and the effort level that the principal chooses to implement (which in turn depends on all the parameters in the model, including the particular functional form of the agent's utility with respect to payments and disutility cost of effort). While explicitly solving for the optimal implemented level of effort is fruitless, we would like to study the qualitative effects of changes in optimism and overconfidence, in order to compare them with our results in the simpler binary-action model.

Although we could modify the proof of Proposition 5 to show that the implemented level of effort increases with agent optimism, such approach does not prove that it also increases with agent overconfidence. I will instead assume that the principal's profit-maximization problem when choosing which effort level to implement is *well behaved*: it has a unique, interior, local and global

maximum. Let e^* denote the effort level that solves the principal's profit maximization problem, at which the marginal revenue from increasing the implemented level of effort equals its marginal cost from the principal's point of view:

$$MR_{e^*} = MC_{e^*}.$$

Note that the marginal revenue of effort is independent of effort level:

$$MR_e = v\left(x_1 - x_0\right).$$

By marginally increasing the implemented level of effort, the additional revenue in the event of project success $(x_1 - x_0)$ will come about with marginally higher probability (note that it depends on v, but not on the agent's beliefs $\tilde{v}$).

The marginal cost of implementing effort, on the other hand, is

$$MC_e = v\left(s_1\left(e\right) - s_0\left(e\right)\right) + \left(q + ve\right)\frac{ds_1\left(e\right)}{de} + \left[1 - \left(q + ve\right)\right]\frac{ds_0\left(e\right)}{de}$$

where

$$\frac{ds_1\left(e\right)}{de} = \frac{1}{u'\left(s_1\left(e\right)\right)}\left[1 - \left(\tilde{q} + \tilde{v}e\right)\right]\frac{c''\left(e\right)}{\tilde{v}}, \text{ and}$$
$$\frac{ds_0\left(e\right)}{de} = -\frac{1}{u'\left(s_0\left(e\right)\right)}\left(\tilde{q} + \tilde{v}e\right)\frac{c''\left(e\right)}{\tilde{v}}.$$

Note, in particular, that the marginal cost of implementing effort does depend crucially on the agent's beliefs. Because marginal revenue is constant in terms of effort, we can guarantee that the principal's profit-maximization will be well behaved if the marginal cost of implementing effort is an increasing function of effort level. For a discussion about the conditions under which this is the case, refer to the Appendix Section A.2.1.

Consider the effect of an increase in agent optimism on the marginal cost of implementing effort, evaluated at the optimal e^*:

$$\frac{\partial MC_{e^*}}{\partial \tilde{q}} = -\frac{c''\left(e^*\right)}{\tilde{v}}\left[\left(q + ve^*\right)\frac{1}{u'\left(s_1\left(e^*\right)\right)} + \left[1 - \left(q + ve^*\right)\right]\frac{1}{u'\left(s_0\left(e^*\right)\right)}\right] < 0.$$

Given that the marginal revenue of implementing any effort level is constant, and that the marginal cost of implementing effort increases with effort level, it follows that **the principal will implement higher effort if dealing with a more optimistic agent**: $\frac{de^*}{d\tilde{q}} > 0$. This result is analogous to its counterpart in the two-action case, summarized in Proposition 5. As a consequence of the wager effect, because a more-optimistic agent prefers higher-powered incentive contracts, it is cheaper for the principal to implement a higher level of effort in the margin.

Consider now the comparable effect of an increase in agent overconfidence:

$$\frac{\partial MC_{e^*}}{\partial \tilde{v}} = -e^* \frac{c''(e^*)}{\tilde{v}} \left[(q + ve^*) \frac{1}{u'(s_1(e^*))} + [1 - (q + ve^*)] \frac{1}{u'(s_0(e^*))} \right] - \frac{1}{\tilde{v}} v \left[(x_1 - s_1(e^*)) - (x_0 - s_0(e^*)) \right] < 0.$$

The first term of the equation above reflects the wager effect. Just as in the case of an increase in agent optimism, it is less costly for the principal to implement higher effort levels. The second term of the equation reflects the incentive effect of overconfidence. The contract $\langle s_1(e^*), s_0(e^*) \rangle$ will implement some effort level greater than e^* following an increase in the agent's overconfidence about the value of effort. This unambiguously benefits the principal as long as $(x_1 - s_1(e^*)) \geq (x_0 - s_0(e^*))$.[15] Implementing a higher level of effort increases the expected revenue of the project. As a consequence of both the wager and the incentive effects of overconfidence, **the principal will implement higher effort if dealing with a more overconfident agent**: $\frac{de^*}{d\tilde{v}} > 0$. This is analogous to the corresponding result in the two-action case, exposed in Proposition 5.

The comparative statics of the power of incentives of the optimal contract with respect to an increase in agent optimism are straightforward. Given that a higher effort level is implemented, and that optimism does not directly affect the incentive structure of the contract if effort was held constant ($\tilde{q}$ is absent from (IC')), it follows that an increase in agent optimism always implies higher-powered incentives. The wager effect implies a reduction in the marginal cost of implementing any effort level, and a higher implemented level of effort drives the optimal contract towards higher-powered incentives in the continuous-action case. This is true even if the agent is only slightly overconfident overall, because the implemented effort level increases with optimism in a continuous fashion.

The comparative statics of the power of incentives of the optimal contract with respect to an increase in agent overconfidence are more subtle, and formally derived in the Appendix (Subsection A.2.2). The incentive effect still pushes towards lower-powered incentives, but both the wager

[15] If the wager effect is so strong that $(x_1 - s_1(e^*)) < (x_0 - s_0(e^*))$, then the principal's profits are higher under project failure than success. If this is the case, the principal faces a tradeoff between the cost savings from increasing the "side-betting" with the agent and a marginally higher probability of losing that wager from implementing higher effort (when the project succeeds). This hints at why we cannot simply extend the envelope theorem approach we used in the proof of Proposition 5 to analyze the effect of an increase in overconfidence (the result in terms of optimism does carry over). Note, however, that in the extreme case in which $(x_1 - s_1(e^*)) < (x_0 - s_0(e^*))$ the principal would have incentives to sabotage the project. If sabotage is feasible, the agent will anticipate this and not accept a contract with higher residual value for the principal if the project fails. In what follows, I maintain the assumption that contracts are restricted to $(x_1 - s_1) \geq (x_0 - s_0)$.

effect and the increase in the implemented level of effort push towards higher-powered incentives. One of the results of the simpler binary-action setting—that there will be some range of slight overconfidence overall such that the power of incentives decreases with overconfidence—carries over to this setting if the agent is sufficiently risk averse and the marginal cost of exerting effort does not increase sharply. Under these conditions, the resulting increase in the implemented level of effort can be achieved with lower power of incentives. The incentive effect dominates the wager effect in terms of the power of incentives, because there is a large increase in the effort level that the initial contract implements; so much so, that the power of incentives is reduced at the same time that a higher optimal effort level is implemented.

Finally, note that the welfare analysis in Subsection 3.3 readily extends to the continuous-action setting; as long as the equilibrium implies an interior solution for the agent's problem, the principal's expected profit increases in both agent optimism and overconfidence.

5. Conclusion

This paper attempts to provide insight into the effects of overconfidence in equilibrium within a moral-hazard framework. It shows that overconfidence may result in optimal (second-best) incentive contracts with lower or higher power of incentives, depending both on the overall level of overconfidence and on the particular type of bias (optimism and overconfidence). It also shows that, in an agency setting, agent overconfidence can affect efficiency as well as distributional considerations.

Because the relationship between overconfidence and power of incentives in our model depends on the level of overall overconfidence, it is not straight-forward to design an empirical test of the effects of overconfidence on power of incentives. If we had access to the details in CEO incentive contracts, a split of a given sample based on a binary measure of overconfidence (e.g. whether or not a CEO is referred to in the press as being overconfident) might be a good approximation to splitting the "significantly overconfident" from the "slightly overconfident" CEOs. If this is the case, our model implies a positive relationship between a continuous measure of overconfidence, like those constructed by Malmendier and Tate (2005), and the power of incentives for the "significantly overconfident" sub-sample (after controlling for industry, size, and other effects), and a negative relationship for the "slightly overconfident" sub-sample.

In terms of the level of effort implemented in equilibrium, on the other hand, the implications are unambiguous: higher optimism or overconfidence implies a higher implemented effort level. In their survey, Cooper et al. (1988) find that entrepreneurs tend to overestimate the probability of success of their enterprise and that they invest many hours in it (more than 60 hours a week

according to many of the respondents). The coincidence of overconfidence with high effort is consistent with the model, whether their overestimation of the probability of success stems from optimism or overconfidence.

The results of the paper suggest that incentive contracts are sensitive to the *kind and level* of overconfidence, not only to the presence of overconfidence per se. This underscores the importance of experimental or field studies that delve into the nuances of overconfidence. Such studies would help our understanding of incentive contracts when entrepreneurs, managers, or employees are overconfident. For instance, if agents tend to be significantly overconfident overall and agent overconfidence is procyclical (as suggested by Gervais and Odean (2001)), our model predicts that fast-paced growth should be followed by more powerful incentive contracts being implemented. In contrast, if agents tend to be only slightly overconfident overall, less powerful incentives would follow. The type of agent bias (optimism or overconfidence) would also affect self-selection results. According to adverse-selection models that allow for overconfidence, the most overconfident agents are attracted to riskier endeavors. This is consistent with the fact that some agents in dangerous jobs do underestimate the probability of a bad outcome, as noted by Akerlof and Dickens (1982). My model implies, however, that different kinds of overconfidence can have conflicting effects in terms of the amount of risk borne by the agent in equilibrium. If this is the case, agents with similar degrees of overall overconfidence might sort themselves into very different positions.

A. Appendix

A.1. Binary Action Space

Proposition 1 *The first-best contract that implements effort level* $e \in \{0,1\}$, $\left\langle s_1^{FBe}, s_0^{FBe} \right\rangle$, *is characterized by the conditions*

$$\frac{\tilde{q}+\tilde{v}e}{1-(\tilde{q}+\tilde{v}e)}\frac{u'\left(s_1^{FBe}\right)}{u'\left(s_0^{FBe}\right)} = \frac{q+ve}{1-(q+ve)}$$

and $(\tilde{q}+\tilde{v}e)\,u\left(s_1^{FBe}\right)+[1-(\tilde{q}+\tilde{v}e)]\,u\left(s_0^{FBe}\right)-ce=\underline{u}$.

Proof. The first-best contract that implements effort level $e \in \{0,1\}$ is the pair of payments which maximizes the principal's expected profit conditional on effort level e being implemented,

$$\mathbb{E}\left[\pi \mid e\right] = (q+ve)\,(x_1-s_1)+[1-(q+ve)]\,(x_0-s_0)\,,$$

subject only to the agent's individual-rationality (IR) constraint,

$$(\tilde{q}+\tilde{v}e)\,u\left(s_1\right)+[1-(\tilde{q}+\tilde{v}e)]\,u\left(s_0\right)-ce \geq \underline{u},$$

since incentive compatibility is not required of the first-best contract.

First, note that (IR) must bind; otherwise, the principal could reduce s_0 or s_1 and thus increase her expected profit. So $(\tilde{q}+\tilde{v}e)\,u\left(s_1^{FBe}\right)+\left[1-(\tilde{q}+\tilde{v}e)\right]u\left(s_0^{FBe}\right)-ce=\underline{u}$.

We can write the Lagrangian

$$\mathcal{L}=(q+ve)\,(x_1-s_1)+\left[1-(q+ve)\right](x_0-s_0)+\lambda\left[(\tilde{q}+\tilde{v}e)\,u\,(s_1)+\left[1-(\tilde{q}+\tilde{v}e)\right]u\,(s_0)-ce-\underline{u}\right].$$

The necessary first-order conditions with respect to s_1 and s_0 yield

$$\frac{\tilde{q}+\tilde{v}e}{1-(\tilde{q}+\tilde{v}e)}\frac{u'\left(s_1^{FBe}\right)}{u'\left(s_0^{FBe}\right)}=\frac{q+ve}{1-(q+ve)}. \tag{A1}$$

By working with the agent's utility levels $\upsilon_x\equiv u\,(s_x)$ and the inverse-utility function $h\equiv u^{-1}$, (Grossman and Hart, 1983, p. 13) show that the principal's problem can be rewritten as an optimization problem of minimizing a convex function subject to (in our case at most two) linear constraints. The first-order condition (A1) above is thus necessary and sufficient for global optimality. Note that our agent utility specification satusfies assumption A1 in (Grossman and Hart, 1983, p. 10).[16] ■

Corollary 1 *If agent preferences are such that* $\frac{1}{u'(s)}\left(-\frac{u''(s)}{u'(s)}\right)$ *is a non-decreasing function, the first-best contract,* $\left\langle s_1^{FBe},s_0^{FBe}\right\rangle$*, exhibits power of incentives that increases in the degree of agent optimism and overconfidence:* $\frac{d}{d\tilde{q}}\left(u\left(s_1^{FBe}\right)-u\left(s_0^{FBe}\right)\right)>0$ *for* $e\in\{0,1\}$ *and* $\frac{d}{d\tilde{v}}\left(u\left(s_1^{FBe}\right)-u\left(s_0^{FBe}\right)\right)>0$ *for* $e=1$.

Proof. The first-best contract $\left\langle s_1^{FBe},s_0^{FBe}\right\rangle$ is characterized in Proposition 1.

We will prove the statement for an optimistic or overall overconfident agent first; the proof for an overall underconfident or pessimistic agent and for one that agrees with the principal's beliefs follow immediately.

If the agent is optimistic or overall overconfident ($\tilde{q}+\tilde{v}e>q+ve$, for $e=0$ or $e=1$, respectively), then optimal risk sharing under disagreement as shown in (A1), given $u''(\cdot)<0$, implies $u\left(s_1^{FBe}\right)>u\left(s_0^{FBe}\right)$.

By taking the total derivative of the binding participation constraint (IR) with respect to $\tilde{q}$, we find:

$$\left(u\left(s_1^{FBe}\right)-u\left(s_0^{FBe}\right)\right)+(\tilde{q}+\tilde{v}e)\frac{du\left(s_1^{FBe}\right)}{d\tilde{q}}+\left[1-(\tilde{q}+\tilde{v}e)\right]\frac{du\left(s_0^{FBe}\right)}{d\tilde{q}}=0.$$

[16] For the same reason, in following proofs the first-order conditions of the principal's problem are necessary and sufficient for global optimality; we would be adding one more linear constraint, that of incentive-compatibility, to the modified problem in second-best implementation.

Given $\left(u\left(s_1^{FBe}\right)-u\left(s_0^{FBe}\right)\right)>0$, it follows immediately that $\frac{du\left(s_1^{FBe}\right)}{d\tilde{q}}<0$ or $\frac{du\left(s_0^{FBe}\right)}{d\tilde{q}}<0$ (or both).

Taking now the total derivative of the optimal risk-sharing rule under disagreement (A1) with respect to $\tilde{q}$, we find:

$$\frac{\tilde{q}+\tilde{\nu}e}{1-(\tilde{q}+\tilde{\nu}e)}\frac{d}{d\tilde{q}}\left(\frac{u'\left(s_1^{FBe}\right)}{u'\left(s_0^{FBe}\right)}\right)+\frac{u'\left(s_1^{FBe}\right)}{u'\left(s_0^{FBe}\right)}\frac{d}{d\tilde{q}}\left(\frac{\tilde{q}+\tilde{\nu}e}{1-(\tilde{q}+\tilde{\nu}e)}\right)=0.$$

Given that both $\frac{\tilde{q}+\tilde{\nu}e}{1-(\tilde{q}+\tilde{\nu}e)}$ and $\frac{u'\left(s_1^{FBe}\right)}{u'\left(s_0^{FBe}\right)}$ are positive, and that $\frac{d}{d\tilde{q}}\left(\frac{\tilde{q}+\tilde{\nu}e}{1-(\tilde{q}+\tilde{\nu}e)}\right)=\frac{1}{[1-(\tilde{q}+\tilde{\nu}e)]^2}>0$, it must be the case that

$$\frac{d}{d\tilde{q}}\left(\frac{u'\left(s_1^{FBe}\right)}{u'\left(s_0^{FBe}\right)}\right)<0. \tag{A2}$$

Assume that $\frac{du\left(s_0^{FBe}\right)}{d\tilde{q}}\geq 0$. Since $\frac{du\left(s_1^{FBe}\right)}{d\tilde{q}}$ and $\frac{du\left(s_0^{FBe}\right)}{d\tilde{q}}$ cannot be both positive, then $\frac{du\left(s_1^{FBe}\right)}{d\tilde{q}}<0$, but then (A2) is violated. Therefore $\frac{du\left(s_0^{FBe}\right)}{d\tilde{q}}<0$.

If $\frac{du\left(s_1^{FBe}\right)}{d\tilde{q}}\geq 0$, then $\frac{d}{d\tilde{q}}\left(u\left(s_1^{FBe}\right)-u\left(s_0^{FBe}\right)\right)>0$ follows immediately.

If both $\frac{du\left(s_1^{FBe}\right)}{d\tilde{q}}<0$ and $\frac{du\left(s_0^{FBe}\right)}{d\tilde{q}}<0$, first note that, taking into account $\frac{du'(s)}{d\tilde{q}}=\frac{u''(s)}{u'(s)}\frac{du(s)}{d\tilde{q}}$, we can write (A2) as

$$\frac{1}{u'\left(s_1^{FBe}\right)}\left(-\frac{u''\left(s_1^{FBe}\right)}{u'\left(s_1^{FBe}\right)}\right)\frac{du\left(s_1^{FBe}\right)}{d\tilde{q}}-\frac{1}{u'\left(s_0^{FBe}\right)}\left(-\frac{u''\left(s_0^{FBe}\right)}{u'\left(s_0^{FBe}\right)}\right)\frac{du\left(s_0^{FBe}\right)}{d\tilde{q}}>0.$$

Given $u\left(s_1^{FBe}\right)>u\left(s_0^{FBe}\right)$, by assumption $\frac{1}{u'\left(s_1^{FBe}\right)}\left(-\frac{u''\left(s_1^{FBe}\right)}{u'\left(s_1^{FBe}\right)}\right)\geq\frac{1}{u'\left(s_0^{FBe}\right)}\left(-\frac{u''\left(s_0^{FBe}\right)}{u'\left(s_0^{FBe}\right)}\right)$, so we can write

$$\frac{1}{u'\left(s_1^{FBe}\right)}\left(-\frac{u''\left(s_1^{FBe}\right)}{u'\left(s_1^{FBe}\right)}\right)\frac{du\left(s_1^{FBe}\right)}{d\tilde{q}}-\frac{1}{u'\left(s_0^{FBe}\right)}\left(-\frac{u''\left(s_0^{FBe}\right)}{u'\left(s_0^{FBe}\right)}\right)\frac{du\left(s_0^{FBe}\right)}{d\tilde{q}}\leq$$
$$\frac{1}{u'\left(s_1^{FBe}\right)}\left(-\frac{u''\left(s_1^{FBe}\right)}{u'\left(s_1^{FBe}\right)}\right)\left(\frac{du\left(s_1^{FBe}\right)}{d\tilde{q}}-\frac{du\left(s_0^{FBe}\right)}{d\tilde{q}}\right).$$

Then $\frac{d}{d\tilde{q}}\left(u\left(s_1^{FBe}\right)-u\left(s_0^{FBe}\right)\right)>0$ follows, since the left-hand side is strictly positive.

If the agent is pessimistic or overall underconfident ($\tilde{q}+\tilde{\nu}e<q+\nu e$), then optimal risk sharing (A2) implies $u\left(s_0^{FBe}\right)>u\left(s_1^{FBe}\right)$. Following the same steps as for the optimistic or overall overconfident case above, we can show that it must be the case that $\frac{du\left(s_1^{FBe}\right)}{d\tilde{q}}>0$. If $\frac{du\left(s_0^{FBe}\right)}{d\tilde{q}}\leq 0$, then the proof is complete. If both $\frac{du\left(s_1^{FBe}\right)}{d\tilde{q}}>0$ and $\frac{du\left(s_0^{FBe}\right)}{d\tilde{q}}>0$, the assumption that $\frac{1}{u'(s)}\left(-\frac{u''(s)}{u'(s)}\right)$ be a non-decreasing function is again sufficient to guarantee $\frac{d}{d\tilde{q}}\left(u\left(s_1^{FBe}\right)-u\left(s_0^{FBe}\right)\right)>0$.

If the agent agrees with the principal's beliefs ($\tilde{q}+\tilde{\nu}e=q+\nu e$), following the same steps we can show that $\frac{du\left(s_1^{FBe}\right)}{d\tilde{q}}>0$ and $\frac{du\left(s_0^{FBe}\right)}{d\tilde{q}}<0$, which implies $\frac{d}{d\tilde{q}}\left(u\left(s_1^{FBe}\right)-u\left(s_0^{FBe}\right)\right)>0$.

The proof that $\frac{d}{d\tilde{\nu}}\left(u\left(s_1^{FBe}\right)-u\left(s_0^{FBe}\right)\right)>0$ for $e=1$ is virtually identical. ■

Proposition 2 *If the agent is slightly overconfident overall, the second-best contract that implements effort is* $\langle \bar{s}_1, \bar{s}_0 \rangle$.

Proof. The second-best contract is the pair of payments which maximizes the principal's expected profit conditional on effort being implemented,

$$\mathbb{E}\left[\pi \mid e = 1\right] = (q + v)\left(x_1 - s_1\right) + \left[1 - (q + v)\right]\left(x_0 - s_0\right),$$

subject to the agent's individual-rationality (IR) constraint,

$$(\tilde{q} + \tilde{v})\, u\left(s_1\right) + \left[1 - (\tilde{q} + \tilde{v})\right] u\left(s_0\right) - c \geq \underline{u},$$

and the agent's incentive-compatibility (IC) constraint,

$$\tilde{v}\left(u\left(s_1\right) - u\left(s_0\right)\right) \geq c.$$

First, note that (IR) must bind; otherwise, the principal could reduce s_0 and thus increase her expected profit without destroying incentives.

We wil show that the (IC) constraint binds if the agent is *slightly overconfident overall*, which implies that the binding (IR) and (IC) constraints define the second-best contract, $\langle \bar{s}_1, \bar{s}_0 \rangle$.

In order to show this, we will show that if the agent is *slightly overconfident overall*, i.e. if

$$\frac{\tilde{q} + \tilde{v}}{1 - (\tilde{q} + \tilde{v})} \frac{u'\left(\bar{s}_1\right)}{u'\left(\bar{s}_0\right)} \leq \frac{q + v}{1 - (q + v)},$$

a change in the offered contract away from $\langle \bar{s}_1, \bar{s}_0 \rangle$ that increases the principal's expected profits and maintains the (IR) constraint as binding necessarily violates the (IC) constraint.

By calculating the total differential of the binding (IR) constraint,

$$(\tilde{q} + \tilde{v})\, u\left(\bar{s}_1\right) + \left[1 - (\tilde{q} + \tilde{v})\right] u\left(\bar{s}_0\right) - c = \underline{u},$$

we find the set of marginal changes in the contract that maintain a binding (IR) constraint, defined by pairs $d\bar{s}_1$ and $d\bar{s}_0$ such that:

$$(\tilde{q} + \tilde{v})\, u'\left(\bar{s}_1\right) d\bar{s}_1 + \left[1 - (\tilde{q} + \tilde{v})\right] u'\left(\bar{s}_0\right) d\bar{s}_0 = 0,$$

which we can write as

$$d\bar{s}_0 = -\frac{\tilde{q} + \tilde{v}}{1 - (\tilde{q} + \tilde{v})} \frac{u'\left(\bar{s}_1\right)}{u'\left(\bar{s}_0\right)} d\bar{s}_1.$$

So in order to maintain a binding (IR) constraint, the change in the contract must either imply an increase in the power of incentives, in which case an increase in s_1 from $\bar{s}_1$ must be accompanied

by a decrease in s_0 with respect to $\bar{s}_0$, or a decrease in the power of incentives, in which case a decrease in s_1 with respect to $\bar{s}_1$ must be accompanied by an increase in s_0 with respect to $\bar{s}_0$.

At the same time we want to restrict the change in the contract to one which increases expected profit for the principal, in which case the total differential of the expected profit for the principal must be positive:

$$-(q+v)\,d\bar{s}_1 - [1-(q+v)]\,d\bar{s}_0 > 0.$$

Substituting for $d\bar{s}_0$ from above, we find that increasing expected profit for the principal requires

$$-(q+v)\,d\bar{s}_1 + [1-(q+v)]\,\frac{\tilde{q}+\tilde{v}}{1-(\tilde{q}+\tilde{v})}\frac{u'(\bar{s}_1)}{u'(\bar{s}_0)}d\bar{s}_1 > 0,$$

or

$$\left(\frac{\tilde{q}+\tilde{v}}{1-(\tilde{q}+\tilde{v})}\frac{u'(\bar{s}_1)}{u'(\bar{s}_0)} - \frac{(q+v)}{[1-(q+v)]}\right)d\bar{s}_1 > 0. \tag{A3}$$

if the agent is *slightly overconfident overall* $\frac{\tilde{q}+\tilde{v}}{1-(\tilde{q}+\tilde{v})}\frac{u'(\bar{s}_1)}{u'(\bar{s}_0)} \leq \frac{q+v}{1-(q+v)}$, which implies that $d\bar{s}_1 < 0$ and $d\bar{s}_0 > 0$. Such a decrease in the power of incentives would violate the (IC) constraint, since it was binding by construction under $\langle \bar{s}_1, \bar{s}_0 \rangle$. ■

Corollary 2 *If the agent is slightly overconfident overall, the second-best contract that implements effort, $\langle \bar{s}_1, \bar{s}_0 \rangle$, exhibits power of incentives that decreases in the degree of agent overconfidence: $\frac{d}{d\tilde{v}}\left(u(\bar{s}_1) - u(\bar{s}_0)\right) < 0$. Agent optimism does not affect the power of incentives: $\frac{d}{d\tilde{q}}\left(u(\bar{s}_1) - u(\bar{s}_0)\right) = 0$.*

Proof. The second-best contract $\langle \bar{s}_1, \bar{s}_0 \rangle$ is characterized, in particular, by the binding (IC) constraint which we can rewrite as

$$\left(u(\bar{s}_1) - u(\bar{s}_0)\right) = \frac{c}{\tilde{v}}.$$

Given $u(\bar{s}_1) - u(\bar{s}_0) = \frac{c}{\tilde{v}} > 0$, it follows immediately that $\frac{d}{d\tilde{v}}\left(u(\bar{s}_1) - u(\bar{s}_0)\right) < 0$ and $\frac{d}{d\tilde{q}}\left(u(\bar{s}_1) - u(\bar{s}_0)\right) = 0$. ■

Proposition 3 *If the agent is significantly overconfident overall, the first-best contract that implements effort, $\langle s_1^*, s_0^* \rangle$, is incentive compatible and thus also the second-best contract that implements effort.*

Proof. If the agent is *significantly overconfident overall*,

$$\frac{\tilde{q}+\tilde{v}}{1-(\tilde{q}+\tilde{v})}\frac{u'(\bar{s}_1)}{u'(\bar{s}_0)} > \frac{q+v}{1-(q+v)}.$$

Refer to condition (A3) above, which defines a change in the contract $\langle \bar{s}_1, \bar{s}_0 \rangle$ that both maintains a binding (IR) constraint and increases expected profit for the principal. If the agent is *significantly*

overconfident overall, then such a change has $d\bar{s}_1 > 0$ and $d\bar{s}_0 < 0$. Such a change increases the power of incentives with respect to $\langle \bar{s}_1, \bar{s}_0 \rangle$, so it (strictly) satisfies incentive compatibility. Given that there exists a change in the $\langle \bar{s}_1, \bar{s}_0 \rangle$ contract that strictly satisfies the (IC) constraint, maintains the (IR) constraint, and increases expected profit for the principal, it must be the case that (IC) is slack at the optimum.

We can thus solve the problem ignoring that constraint. We solved that problem in Proposition 1, so the first-best contract that implements effort, $\langle s_1^*, s_0^* \rangle$, is incentive compatible and thus also the second-best contract that implements effort.

Alternatively, one could directly prove that $\langle s_1^*, s_0^* \rangle$ is strictly incentive compatible if the agent is significantly overconfident overall. Note that both $\langle \bar{s}_1, \bar{s}_0 \rangle$ and $\langle s_1^*, s_0^* \rangle$ satisfy the binding (IR) constraint, that $\langle \bar{s}_1, \bar{s}_0 \rangle$ satisfies the binding (IC) constraint, and that $u'' < 0$. If the agent is significantly overconfident overall, $\frac{u'(\bar{s}_1)}{u'(\bar{s}_0)} > \frac{u'(s_1^*)}{u'(s_0^*)}$, and therefore $\tilde{v}\left(u\left(s_1^*\right) - u\left(s_0^*\right)\right) > c$. ■

Corollary 3 *If the agent is significantly overconfident overall, and agent preferences are such that* $\frac{1}{u'(s)}\left(-\frac{u''(s)}{u'(s)}\right)$ *is a non-decreasing function, the second-best contract that implements effort,* $\langle s_1^*, s_0^* \rangle$*, exhibits power of incentives that increases in the degree of agent optimism and overconfidence:* $\frac{d}{d\tilde{q}}\left(u\left(s_1^*\right) - u\left(s_0^*\right)\right) > 0$ *and* $\frac{d}{d\tilde{v}}\left(u\left(s_1^*\right) - u\left(s_0^*\right)\right) > 0$.

Proof. Apply Corollary 1, since by Proposition 3 $\langle s_1^*, s_0^* \rangle \equiv \langle s_1^{FBe}, s_0^{FBe} \rangle$ for $e = 1$ if the agent is significantly overconfident overall. ■

Proposition 4 *If the agent is slightly optimistic, the first-best contract that implements no effort,* $\langle s_{1*}, s_{0*} \rangle$*, is incentive compatible and thus also the second-best contract that implements effort. If the agent is significantly optimistic, the second-best contract that implements no effort is* $\langle \underline{s}_1, \underline{s}_0 \rangle$.

Proof. The proof is analogous to that of Propositions 3 and 2, respectively, for a slightly or significantly optimistic agent. ■

Corollary 4 *If the agent is slightly optimistic, and agent preferences are such that* $\frac{1}{u'(s)}\left(-\frac{u''(s)}{u'(s)}\right)$ *is a non-decreasing function, the second-best contract that implements no effort,* $\langle s_{1*}, s_{0*} \rangle$*, exhibits power of incentives that increases in the degree of agent optimism:* $\frac{d}{d\tilde{q}}\left(u\left(s_{1*}\right) - u\left(s_{0*}\right)\right) > 0$*. Agent overconfidence does not affect the power of incentives:* $\frac{d}{d\tilde{v}}\left(u\left(s_{1*}\right) - u\left(s_{0*}\right)\right) = 0$.

Proof. Apply Corollary 1, since by Proposition 4 $\langle s_{1*}, s_{0*} \rangle \equiv \langle s_1^{FBe}, s_0^{FBe} \rangle$ for $e = 0$ if the agent is slightly optimistic. ■

Corollary 5 *If the agent is significantly optimistic, the second-best contract that implements no*

effort, $\langle \underline{s}_1, \underline{s}_0 \rangle$*, exhibits power of incentives that decreases in the degree of agent overconfidence:* $\frac{d}{d\tilde{v}} \left(u\left(\underline{s}_1\right) - u\left(\underline{s}_0\right)\right) < 0$*. Agent optimism does not affect the power of incentives:* $\frac{d}{d\tilde{q}} \left(u\left(\underline{s}_1\right) - u\left(\underline{s}_0\right)\right) = 0$*.*

Proof. Analogous to the proof of Corollary 2. ■

Proposition 5 *Holding other parameters constant, if effort is implemented given agent's beliefs* $(\tilde{q}, \tilde{v})$*, then effort will be implemented when his beliefs are* $(\tilde{q}, \tilde{v}')$ *for any* $\tilde{v}' \geq \tilde{v}$ *or* $(\tilde{q}', \tilde{v})$ *for any* $\tilde{q}' \geq \tilde{q}$*.*

Proof. We can write the overall problem for the principal as:

$$\max_{e \in \{0,1\}, s_1, s_0} \mathbb{E}\left[\pi\right] = \left(q + ve\right)\left(x_1 - s_1\right) + \left[1 - \left(q + ve\right)\right]\left(x_0 - s_0\right),$$

subject to the agent's individual-rationality constraint,

$$\left(\tilde{q} + \tilde{v}e\right) u\left(s_1\right) + \left[1 - \left(\tilde{q} + \tilde{v}e\right)\right] u\left(s_0\right) - ce \geq \underline{u},$$

and to the agent's incentive-compatibility constraint:

$$e \in \arg\max_{\hat{e} \in \{0,1\}} \left(\tilde{q} + \tilde{v}\hat{e}\right) u\left(s_1\right) + \left[1 - \left(\tilde{q} + \tilde{v}\hat{e}\right)\right] u\left(s_0\right) - c\hat{e}.$$

Let π^* denote the value function of this problem. We followed Grossman and Hart (1983), breaking up this problem into two stages: finding the second-best way to implement $e = 0$ and $e = 1$, and then comparing the principal's expected profit in these two cases. Let π_e^* denote the principal's expected profits under second-best implementation of effort level $e \in \{0, 1\}$:

$$\begin{gathered} \pi_0^* \equiv \max_{s_1, s_0} q\left(x_1 - s_1\right) + \left[1 - q\right]\left(x_0 - s_0\right) \\ \text{s.t. } \tilde{q}u\left(s_1\right) + \left[1 - \tilde{q}\right] u\left(s_0\right) - \underline{u} \geq 0 \\ \text{and } c - \tilde{v}\left(u\left(s_1\right) - u\left(s_0\right)\right) \geq 0. \end{gathered}$$

and

$$\begin{gathered} \pi_1^* \equiv \max_{s_1, s_0} \left(q + v\right)\left(x_1 - s_1\right) + \left[1 - \left(q + v\right)\right]\left(x_0 - s_0\right) \\ \text{s.t. } \left(\tilde{q} + \tilde{v}\right) u\left(s_1\right) + \left[1 - \left(\tilde{q} + \tilde{v}\right)\right] u\left(s_0\right) - c - \underline{u} \geq 0 \\ \text{and } \tilde{v}\left(u\left(s_1\right) - u\left(s_0\right)\right) - c \geq 0. \end{gathered}$$

The principal will choose to implement effort in equilibrium whenever $\pi_1^* \geq \pi_0^*$, and to implement no effort otherwise. Note that we can apply the envelope theorem to show the first part of our proposition, recalling that the individual-rationality constraint is always binding:

$$\frac{\partial \pi_1^*}{\partial \tilde{v}} = \left(\lambda_1 + \mu_1\right)\left(u\left(s_1\right) - u\left(s_0\right)\right) \geq \left(\lambda_1 + \mu_1\right)\frac{c}{\tilde{v}} > 0,$$

where $\lambda_1 > 0$ and $\mu_1 \geq 0$ are the Lagrange multipliers with respect to the (IR) and (IC) constraints, respectively, when effort is implemented, while

$$\frac{\partial \pi_0^*}{\partial \tilde{v}} = -\mu_0 \left(u\left(s_1\right) - u\left(s_0\right) \right) \leq 0,$$

where $\mu_0 \geq 0$ is the Lagrange multiplier with respect to the incentive-compatibility constraint when no effort is implemented. Leaving all other parameters unchanged, this means that if $\pi_1^*\left(\tilde{q}, \tilde{v}\right) \geq \pi_0^*\left(\tilde{q}, \tilde{v}\right)$, then $\pi_1^*\left(\tilde{q}, \tilde{v}'\right) \geq \pi_0^*\left(\tilde{q}, \tilde{v}'\right)$ for any $\tilde{v}' \geq \tilde{v}$, which completes the first part of our proof.

Following the same logic, we find that

$$\frac{\partial \pi_e^*}{\partial \tilde{q}} = \lambda_e \left(u\left(s_1\right) - u\left(s_0\right) \right) \text{ for } e \in \{0, 1\}.$$

Note that $u\left(s_1\right) - u\left(s_0\right) \geq \frac{c}{\tilde{v}}$ if $e = 1$, while $u\left(s_1\right) - u\left(s_0\right) \leq \frac{c}{\tilde{v}}$ if $e = 0$, which would complete our proof if λ_e was a constant (alas, it is not).

We can, however, exploit the duality of this problem. Note that the solution to the principal's problem (above) is also the solution to

$$\max_{e \in \{0,1\}, s_1, s_0} \widetilde{\mathbb{E}}\left[U\right] = \left(\tilde{q} + \tilde{v}e\right) u\left(s_1\right) + \left[1 - \left(\tilde{q} + \tilde{v}e\right)\right] u\left(s_0\right) - ce$$

subject to the agent's incentive-compatibility constraint,

$$e \in \arg\max_{\hat{e} \in \{0,1\}} \left(\tilde{q} + \tilde{v}\hat{e}\right) u\left(s_1\right) + \left[1 - \left(\tilde{q} + \tilde{v}\hat{e}\right)\right] u\left(s_0\right) - c\hat{e},$$

and the constraint

$$\left(q + ve\right)\left(x_1 - s_1\right) + \left[1 - \left(q + ve\right)\right]\left(x_0 - s_0\right) \geq \pi^*.$$

Let U^* denote the value function of this problem (analogous to π^*; note that $U^* = \underline{u}$ when evaluated at the solution to the principal's problem). The duality can be easily shown by contradiction, given that both principal and agent strictly prefer more money to less:

Let e^*, s_1^*, s_0^* denote the solution to the principal's problem. Assume, however, that it is not a solution to the agent's problem as stated, i.e., assume there is some e', s_1', s_0' such that

$$\left(q + ve'\right)\left(x_1 - s_1'\right) + \left[1 - \left(q + ve'\right)\right]\left(x_0 - s_0'\right) \geq \pi^*,$$

$$e' \in \arg\max_{\hat{e} \in \{0,1\}} \left(\tilde{q} + \tilde{v}\hat{e}\right) u\left(s_1'\right) + \left[1 - \left(\tilde{q} + \tilde{v}\hat{e}\right)\right] u\left(s_0'\right) - c\hat{e},$$

and

$$\left(\tilde{q} + \tilde{v}e'\right) u\left(s_1'\right) + \left[1 - \left(\tilde{q} + \tilde{v}e'\right)\right] u\left(s_0'\right) - ce' > \underline{u}.$$

This means that, in the principal's problem, she could have offered some slightly less expensive contract $\langle s_1'', s_0'' \rangle$ with $s_1'' < s_1'$ and $s_0'' < s_0'$ such that $\left(u\left(s_1''\right) - u\left(s_0''\right)\right) = \left(u\left(s_1'\right) - u\left(s_0'\right)\right)$, which

implements effort level e' and also satisfies the individual-rationality constraint, thus increasing her expected profit.

Applying the envelope theorem to this dual specification of the problem yields

$$\frac{\partial U^*}{\partial \tilde{q}} = \left(u\left(s_1\right) - u\left(s_0\right)\right).$$

Given $u\left(s_1\right) - u\left(s_0\right) \geq \frac{c}{\tilde{v}}$ if $e = 1$, while $u\left(s_1\right) - u\left(s_0\right) \leq \frac{c}{\tilde{v}}$ if $e = 0$, this means $\frac{\partial U_1^*}{\partial \tilde{q}} \geq \frac{\partial U_0^*}{\partial \tilde{q}}$ (notation analogous to π_1^* and π_0^*). This implies that if $U_1^*\left(\tilde{q}, \tilde{v}\right) \geq U_0^*\left(\tilde{q}, \tilde{v}\right)$, then $U_1^*\left(\tilde{q}', \tilde{v}\right) \geq U_0^*\left(\tilde{q}', \tilde{v}\right)$ for any $\tilde{q}' \geq \tilde{q}$. Given the duality of the problem, it follows that if $\pi_1^*\left(\tilde{q}, \tilde{v}\right) \geq \pi_0^*\left(\tilde{q}, \tilde{v}\right)$, then $\pi_1^*\left(\tilde{q}', \tilde{v}\right) \geq \pi_0^*\left(\tilde{q}', \tilde{v}\right)$ for any $\tilde{q}' \geq \tilde{q}$.[17] ■

Proposition 6 *If effort is implemented, the principal's expected profit increases in both the agent's level of optimism and overconfidence. If no effort is implemented, the principal's expected profit increases in the agent's level of optimism or pessimism, for an optimistic or a pessimistic agent, respectively, and it decreases in agent overconfidence if the agent is significantly optimistic.*

Proof. Refer to the proof of Proposition 5 above. We showed that

$$\frac{\partial \pi_1^*}{\partial \tilde{v}} = \left(\lambda_1 + \mu_1\right)\left(u\left(s_1\right) - u\left(s_0\right)\right) > 0,$$

$$\frac{\partial \pi_0^*}{\partial \tilde{v}} = -\mu_0\left(u\left(s_1\right) - u\left(s_0\right)\right) \leq 0,$$

and that

$$\frac{\partial \pi_e^*}{\partial \tilde{q}} = \lambda_e\left(u\left(s_1\right) - u\left(s_0\right)\right) \text{ for } e \in \{0, 1\}.$$

In proving Propositions 1, 2, 3, and 4, we showed that the (IR) constraint is always binding, which implies $\lambda_e > 0$. If effort is implemented, we know that the (IC) constraint is satisfied, so $\left(u\left(s_1\right) - u\left(s_0\right)\right) \geq \frac{c}{\tilde{v}} > 0$. This implies

$$\frac{\partial \pi_1^*}{\partial \tilde{q}} > 0.$$

If no effort is implemented, we know from Proposition 4 that $\left(u\left(s_1\right) - u\left(s_0\right)\right) > 0$ in the case of agent optimism ($\tilde{q} > q$), so

$$\frac{\partial \pi_0^*}{\partial \tilde{q}} > 0 \text{ if the agent is optimistic.}$$

[17] The duality of the problem is also what makes the results in the setting of a principal making a take-it-or-leave-it offer to the agent, studied in the final version of this paper, qualitatively very similar to the results in the setting of two or more principals competing in offering contracts to an agent, studied in the working paper version (de la Rosa (2007)).

If no effort is implemented we also know from Proposition 4 that $(u(s_1) - u(s_0)) < 0$ in the case of agent pesimism ($\tilde{q} < q$), so

$$\frac{\partial \pi_0^*}{\partial \tilde{q}} < 0 \text{ if the agent is pessimistic,}$$

thus an increase in agent pessimism (i.e. a decrease in $\tilde{q}$) increases the principal's expected profit.

Finally, if no effort is implemented and the agent is significantly optimistic, the second-best contract, $\langle \underline{s}_1, \underline{s}_0 \rangle$ is defined by the binding individual-rationality constraint and the binding incentive-compatibility constraint. Since the incentive-compatibility constraint binds, it follows that $\mu_0 > 0$ and

$$\frac{\partial \pi_0^*}{\partial \tilde{v}} = -\mu_0 (u(s_1) - u(s_0)) < 0.$$

If the agent is slightly optimistic or pessimistic, the incentive-compatibility constraint for implementing no effort is slack, so $\mu_0 = 0$ and

$$\frac{\partial \pi_0^*}{\partial \tilde{v}} = -\mu_0 (u(s_1) - u(s_0)) = 0.$$

■

A.2. Continuous Action Space

A.2.1. Conditions for the principal's profit-maximization problem to be well behaved

Recall that we can write the principal's profit-maximization problem as:

$$\max_{e \in [0,1]} (q + ve)(x_1 - s_1(e)) + [1 - (q + ve)](x_0 - s_0(e)).$$

subject to

$$u(s_1(e)) = \underline{u} + c(e) + [1 - (\tilde{q} + \tilde{v}e)] \frac{c'(e)}{\tilde{v}}$$
$$u(s_0(e)) = \underline{u} + c(e) - (\tilde{q} + \tilde{v}e) \frac{c'(e)}{\tilde{v}}.$$

Let e^* denote the solution to the first-order condition of this problem, the level of effort at which the principal's marginal revenue equals her marginal cost of implementing effort:

$$MR_{e^*} = MC_{e^*}.$$

Recall that

$$MR_e = v(x_1 - x_0),$$

and

$$MC_e = v(s_1(e) - s_0(e)) + (q + ve) \frac{ds_1(e)}{de} + [1 - (q + ve)] \frac{ds_0(e)}{de},$$

where

$$\frac{ds_1(e)}{de} = \frac{1}{u'(s_1(e))}\left[1-(\tilde{q}+\tilde{v}e)\right]\frac{c''(e)}{\tilde{v}}$$
$$\frac{ds_0(e)}{de} = -\frac{1}{u'(s_0(e))}(\tilde{q}+\tilde{v}e)\frac{c''(e)}{\tilde{v}}.$$

This problem will have a unique, interior, local and global maximum if the marginal cost of implementing effort is strictly increasing in implemented level of effort. The change in the marginal cost of increasing effort is:

$$\frac{dMC_e}{de} = v\left(\frac{ds_1(e)}{de}-\frac{ds_0(e)}{de}\right)+(q+ve)\frac{d^2s_1(e)}{de^2}+\left[1-(q+ve)\right]\frac{d^2s_0(e)}{de^2}.$$

The first component of this expression is positive, since $\frac{ds_1(e)}{de} > 0$ and $\frac{ds_0(e)}{de} < 0$. Assuming for tractability that $c''(e) = k$, a constant, the second and third components of the expression above are:

$$\frac{d^2s_1(e)}{de^2} = \frac{k}{u'(s_1(e))}\left[-\frac{u''(s_1(e))}{u'(s_1(e))}\frac{\left[1-(\tilde{q}+\tilde{v}e)\right]^2}{\tilde{v}^2}\frac{k}{u'(s_1(e))}-1\right],$$
$$\frac{d^2s_0(e)}{de^2} = \frac{k}{u'(s_0(e))}\left[-\frac{u''(s_0(e))}{u'(s_0(e))}\frac{(\tilde{q}+\tilde{v}e)^2}{\tilde{v}^2}\frac{k}{u'(s_0(e))}-1\right].$$

If both of these components are positive, it follows that the marginal cost of implementing effort will be strictly increasing in effort level. This will be the case if:

- $k\,(=c''(e))$ is large enough. If the cost to the agent of choosing higher levels of effort is convex enough, then the cost to the principal of implementing higher levels of effort will be convex as well.
- the agent is sufficiently risk averse. A large coefficient of absolute risk aversion $-\frac{u''(s_x)}{u'(s_x)}$ also makes it increasingly costly to implement higher effort, since the principal must compensate the agent for the higher risk he must bear as higher levels of effort are implemented.
- the agent is wealthy. It is increasingly costly for the principal to power up incentives and implement higher levels of effort when changes in the payments have little effect on the agent's utility level. When the agent is wealthy, his marginal utility $u'(s_x)$ is relatively low.

A.2.2. The power of incentives and agent overconfidence

Recall that the solution to the agent's problem yields (IC'), which we can write as

$$\left[u(s_1(e^*))-u(s_0(e^*))\right] = \frac{c'(e^*)}{\tilde{v}}.$$

The change in the power of incentives of the equilibrium contract is thus

$$\frac{d\left[u\left(s_1\left(e^*\right)\right)-u\left(s_0\left(e^*\right)\right)\right]}{d\tilde{v}}=\frac{c''\left(e^*\right)}{\tilde{v}}\frac{de^*}{d\tilde{v}}-\frac{c'\left(e^*\right)}{\tilde{v}^2}.$$

Recall, as well, that the solution to the principal's problem e^* is such that

$$MR_{e^*}=MC_{e^*},$$

or

$$v\left(x_1-x_0\right)=v\left(s_1-s_0\right)+\left(q+ve^*\right)\frac{ds_1\left(e^*\right)}{de}+\left[1-\left(q+ve^*\right)\right]\frac{ds_0\left(e^*\right)}{de}.$$

Again, assume that $c''\left(e\right)=k$, a constant. Taking the total derivative of the equation above with respect to $\tilde{v}$ yields

$$\begin{aligned}0=&\frac{de^*}{d\tilde{v}}\left(\left\{\frac{\left[1-\left(\tilde{q}+\tilde{v}e^*\right)\right]}{u'\left(s_1\left(e^*\right)\right)}+\frac{\left(\tilde{q}+\tilde{v}e^*\right)}{u'\left(s_0\left(e^*\right)\right)}\right\}v+\left\{\frac{\left[1-\left(\tilde{q}+\tilde{v}e^*\right)\right]}{u'\left(u\left(s_1\left(e^*\right)\right)\right)}+\frac{\left(\tilde{q}+\tilde{v}e^*\right)}{u'\left(u\left(s_0\left(e^*\right)\right)\right)}\right\}v\right.\\&-\left\{\frac{\left(q+ve^*\right)}{u'\left(u\left(s_1\left(e^*\right)\right)\right)}+\frac{\left[1-\left(q+ve^*\right)\right]}{u'\left(u\left(s_0\left(e^*\right)\right)\right)}\right\}\tilde{v}\\&+\frac{k}{\tilde{v}}\left\{-\frac{u''\left(u\left(s_1\left(e^*\right)\right)\right)}{u'\left(u\left(s_1\left(e^*\right)\right)\right)}\frac{\left(q+ve^*\right)\left[1-\left(\tilde{q}+\tilde{v}e^*\right)\right]^2}{u'\left(u\left(s_1\left(e^*\right)\right)\right)}\right.\\&\left.\left.-\frac{u''\left(u\left(s_0\left(e^*\right)\right)\right)}{u'\left(u\left(s_0\left(e^*\right)\right)\right)}\frac{\left[1-\left(q+ve^*\right)\right]\left(\tilde{q}+\tilde{v}e^*\right)^2}{u'\left(u\left(s_0\left(e^*\right)\right)\right)}\right\}\right)\\&-\frac{1}{\tilde{v}}\left\{\frac{\left(q+ve^*\right)\left[1-\left(\tilde{q}+\tilde{v}e^*\right)\right]}{u'\left(u\left(s_1\left(e^*\right)\right)\right)}+\frac{\left[1-\left(q+ve^*\right)\right]\left(\tilde{q}+\tilde{v}e^*\right)}{u'\left(u\left(s_0\left(e^*\right)\right)\right)}\right\}\\&-e^*\left\{\frac{\left(q+ve^*\right)}{u'\left(u\left(s_1\left(e^*\right)\right)\right)}+\frac{\left[1-\left(q+ve^*\right)\right]}{u'\left(u\left(s_0\left(e^*\right)\right)\right)}\right\}.\end{aligned}$$

Given that we are interested in the effect of overconfidence on the power of incentives when the agent is slightly overconfident, we will evaluate the change in the power of incentives of the equilibrium contract at the point that principal and agent agree in their beliefs (i.e. no overconfidence):

$$\left.\frac{d\left[u\left(s_1\left(e^*\right)\right)-u\left(s_0\left(e^*\right)\right)\right]}{d\tilde{v}}\right|_{\tilde{v}=v,\tilde{q}=q}=\frac{k}{v}\frac{de^*}{d\tilde{v}}-\frac{c'\left(e^*\right)}{v^2}$$

$$\begin{aligned}=&\left[\left(\left\{\frac{\left[1-\left(q+ve^*\right)\right]}{u'\left(s_1\left(e^*\right)\right)}-\frac{\left(q+ve^*\right)}{u'\left(s_0\left(e^*\right)\right)}\right\}+\left\{\frac{1}{\beta_1}-\frac{1}{\beta_0}\right\}\right)v\left(\frac{v}{k}\right)\right.\\&\left.+\left\{-\frac{u''\left(u\left(s_1\left(e^*\right)\right)\right)}{u'\left(u\left(s_1\left(e^*\right)\right)\right)}\frac{\left[1-\left(q+ve^*\right)\right]}{\beta_1}-\frac{u''\left(u\left(s_0\left(e^*\right)\right)\right)}{u'\left(u\left(s_0\left(e^*\right)\right)\right)}\frac{\left(q+ve^*\right)}{\beta_0}\right\}\alpha\right]^{-1}\\&\cdot\left[\left\{\frac{1}{\beta_1}-\frac{1}{\beta_0}\right\}\frac{\alpha}{v}+\left\{\frac{\left(q+ve^*\right)}{\beta_1}-\frac{\left[1-\left(q+ve^*\right)\right]}{\beta_0}\right\}e^*\right]\\&-\frac{c'\left(e^*\right)}{v^2}\end{aligned}$$

where

$$\alpha = (q + ve^*)\left[1 - (q + ve^*)\right], \beta_1 = u'\left(u\left(s_1\left(e^*\right)\right)\right), \beta_0 = u'\left(u\left(s_0\left(e^*\right)\right)\right).$$

If $\frac{d[u(s_1(e^*))-u(s_0(e^*))]}{d\tilde{v}}\Big|_{\tilde{v}=v,\tilde{q}=q} < 0$, then we can say that the power of incentives decreases, following an increase in agent overconfidence, for slight levels of overall overconfidence. Note that the second term of the expression above, $-\frac{c'(e^*)}{v^2} < 0$, will determine the sign of the whole expression if the first term is small enough. As $k \to 0$ and as $-\frac{u''(u(s_x))}{u'(u(s_x))} \to \infty$, the first term converges to zero. So we can guarantee that the result of decreasing power of incentives over some range of overconfidence obtained in the binary-action space setting will carry over to the continuous-action space setting if the increase in the marginal cost of effort is sufficiently low (as measured by $k = c''\left(e^*\right)$), or if the agent is sufficiently risk averse (as measured by $-\frac{u''(u(s_x))}{u'(u(s_x))}$, which slightly resembles the coefficient of absolute risk aversion).

References

Adrian, T. and Westerfield, M. M. 2009. Disagreement and learning in a dynamic contracting model. Review of Financial Studies, 22(10):3873–3906.

Akerlof, G. A. and Dickens, W. T. 1982. The economic consequences of cognitive dissonance. American Economic Review, 72(3):307–319.

Bénabou, R. and Tirole, J. 2002. Self-confidence and personal motivation. Quarterly Journal of Economics, 117(3):871–915.

Bernardo, A. E. and Welch, I. 2001. On the evolution of overconfidence and entrepreneurs. Journal of Economics & Management Strategy, 10(3):301–330.

Borch, K. 1962. Equilibrium in a reinsurance market. Econometrica, 30(3):424–444.

Camerer, C. and Lovallo, D. 1999. Overconfidence and excess entry: An experimental approach. American Economic Review, 89(1):306–318.

Cooper, A. C., Woo, C. Y., and Dunkelberg, W. C. 1988. Entrepreneurs' perceived chances for success. Journal of Business Venturing, 3(2):97–108.

de la Rosa, L. E. 2007. Overconfidence and moral hazard. Working Paper No. 24, Danish Centre for Accounting and Finance. (unpubl.)

de Meza, D. and Southey, C. 1996. The borrower's curse: Optimism, finance and entrepreneurship. The Economic Journal, 106(435):375–386.

Eyster, E. and Rabin, M. 2005. Cursed equilibrium. Econometrica, 73(5):1623–1672.

Gervais, S. and Goldstein, I. 2007. The positive effects of biased self-perceptions in firms. Review of Finance, 11(3):453–496.

Gervais, S., Heaton, J., and Odean, T. 2009. Overconfidence, compensation contracts, and labor markets. mimeo, Fuqua School of Business, Duke University. (unpubl.)

Gervais, S. and Odean, T. 2001. Learning to be overconfident. Review of Financial Studies, 14(1):1–27.

Goel, A. M. and Thakor, A. V. 2008. Overconfidence, CEO selection, and corporate governance. The Journal of Finance, 63(6):2737–2784.

Grossman, S. J. and Hart, O. D. 1983. An analysis of the principal-agent problem. Econometrica, 51(1):7–45.

Holmström, B. 1979. Moral hazard and observability. The Bell Journal of Economics, 10(1):74–91.

Hoorens, V. 1993. Self-enhancement and superiority biases in social comparison. European Review of Social Psychology, 4(1):113–139.

Koufopoulos, K. 2008. Asymmetric information, heterogeneity in risk perceptions and insurance: An explanation to a puzzle. mimeo, Warwick Business School. (unpubl.)

Laffont, J.-J. and Martimort, D. 2002. The Theory of Incentives: The Principal-Agent Model. Princeton University Press, 41 William Street, Princeton, New Jersey 08540.

Larwood, L. and Whittaker, W. 1977. Managerial myopia: Self-serving biases in organizational planning. Journal of Applied Psychology, 62(2):194–198.

Malmendier, U. and Tate, G. 2005. CEO overconfidence and corporate investment. Journal of Finance, 60(6):2661–2700.

Manove, M. and Padilla, A. J. 1999. Banking (conservatively) with optimists. The RAND Journal of Economics, 30(2):324–350.

Maskin, E. and Tirole, J. 1990. The principal-agent relationship with an informed principal: The case of private values. Econometrica, 58(2):379–409.

Maskin, E. and Tirole, J. 1992. The principal-agent relationship with an informed principal, II: Common values. Econometrica, 60(1):1–42.

Moore, D. A. and Healy, P. J. 2008. The trouble with overconfidence. Psychological Review, 115(2):502–517.

Morris, S. 1995. The common prior assumption in economic theory. Economics and Philosophy, 11(2):227–253.

Prendergast, C. 1999. The provision of incentives in firms. Journal of Economic Literature, 37(1):7–63.

Santos-Pinto, L. 2008. Positive self-image and incentives in organizations. The Economic Journal, 118:1315–1332.

Taylor, S. E. and Brown, J. D. 1988. Illusion and well-being: A social psychological perspective on mental health. Psychological Bulletin, 103(2):193–210.

Van den Steen, E. 2004. Rational overoptimism (and other biases). The American Economic Review, 94(4):1141–1151.

Villeneuve, B. 2000. The consequences for a monopolistic insurance firm of evaluating risk better than customers: The adverse selection hypothesis reversed. The Geneva Papers on Risk and Insurance Theory, 25(1):65–79.

Weinstein, N. D. 1980. Unrealistic optimism about future life events. Journal of Personality and Social Psychology, 39(5):806–820.

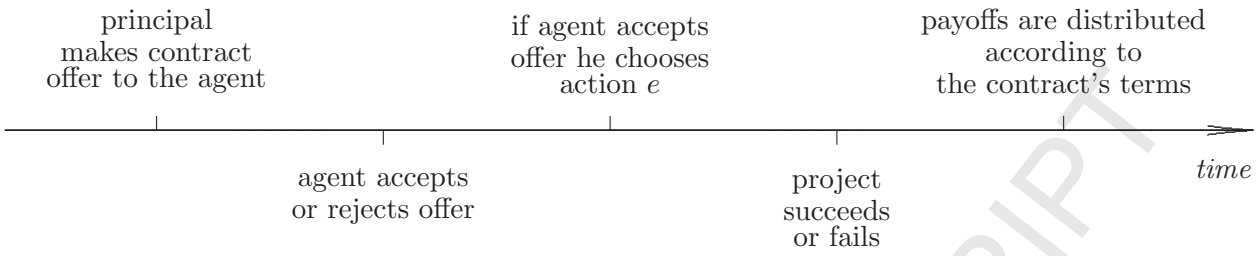

Figure 1: Timing of the model

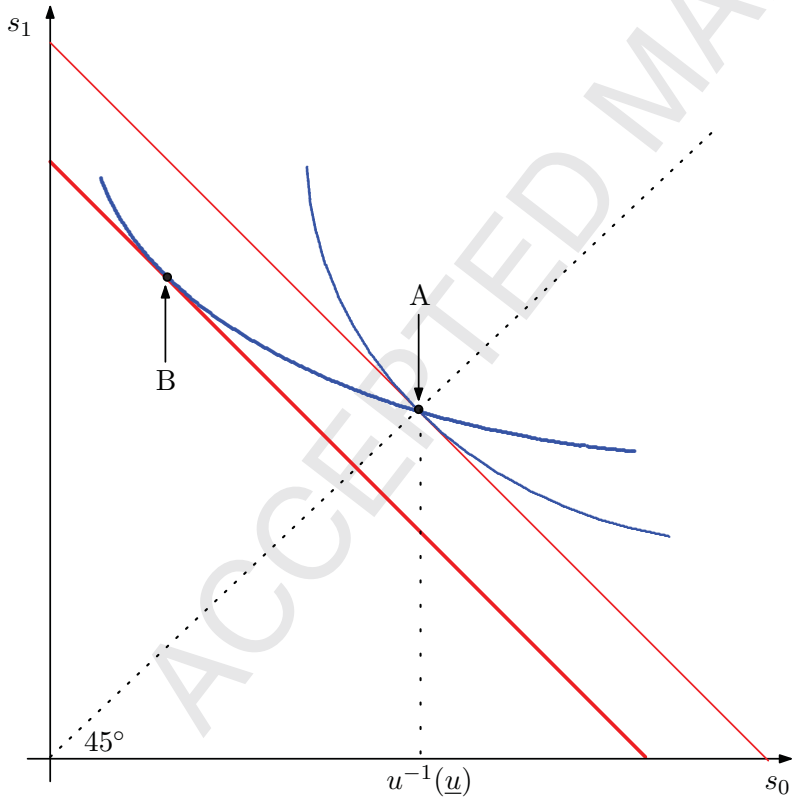

Figure 2: The first-best contract that implements no effort. Given homogeneous beliefs, the risk-averse agent (with indifference curves that are convex to the origin) is fully insured (A). An optimistic agent is offered a risky contract with higher success-contingent pay (B). The risk-neutral principal's iso-expected-profit curves are straight lines.

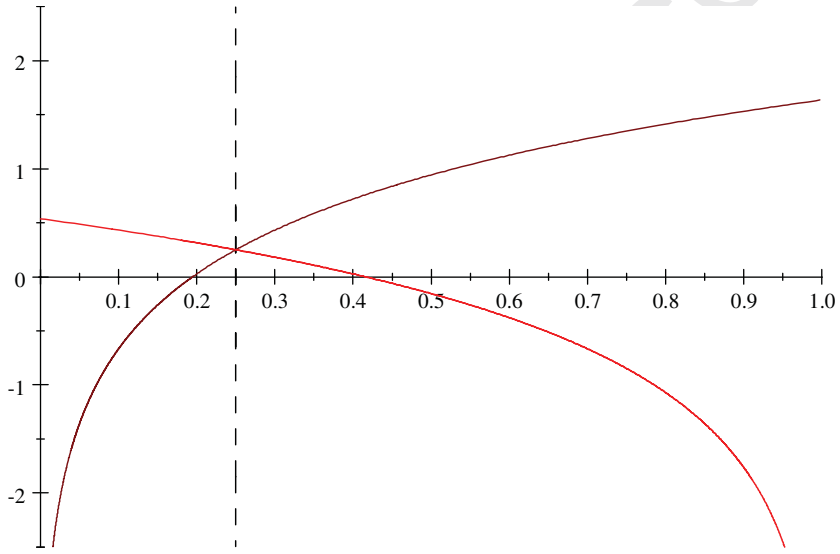

Figure 3: The *wager effect* of overconfidence. We graph s_{1*} (in dark red) and s_{0*} (in light red) as we allow $\tilde{q}$ to vary. For this graph, we use $u(s) = 1 - e^{-s}$, $c = 0.1$, $\underline{u} = 1 - e^{-0.25}$, and $q = 0.25$. Thus, $\tilde{q} = 0.25$ is the case of homogeneous beliefs, $\tilde{q} > 0.25$ means agent optimism, and $\tilde{q} < 0.25$ agent pessimism. As disagreement increases, the agent bears more risk given optimal wagering or "side-betting."

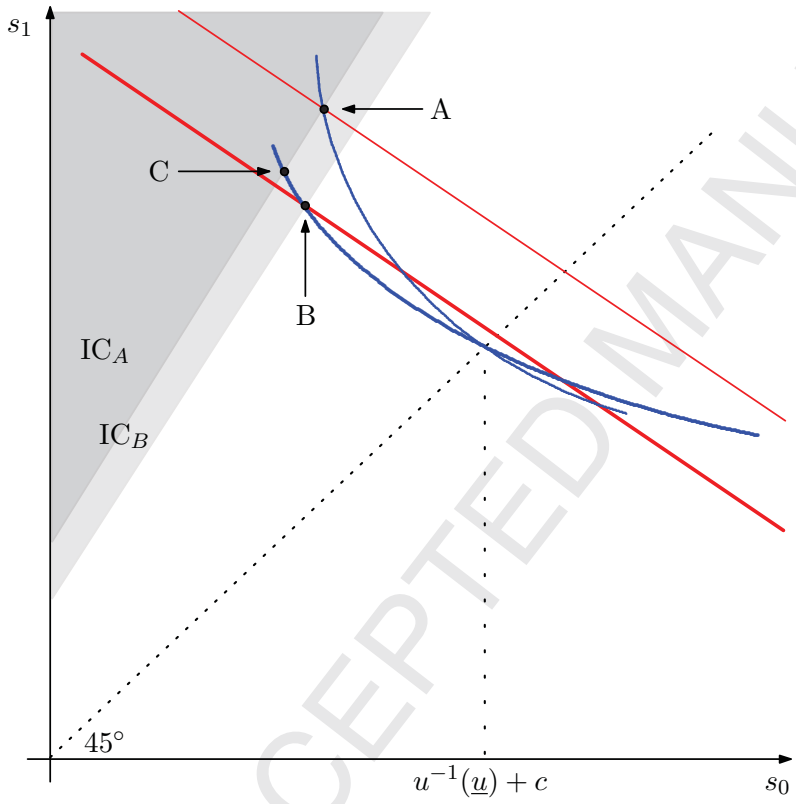

Figure 4: The second-best contract that implements effort (slight overconfidence overall). Given homogeneous beliefs, agent bears just enough risk for the contract to be incentive-compatible (A). A slightly overconfident agent also bears just enough risk to ensure incentive compatibility (B). Slight overconfidence has two effects: it expands the set of incentive-compatible contracts (the set of (s_1, s_0) such that $\tilde{v}\left(u\left(s_1\right) - u\left(s_0\right)\right) \geq c$) and it slackens the participation constraint. For comparison, the second-best contract that implements effort with an optimistic agent who is not overconfident (C) is shown; optimism does not affect the set of incentive-compatible contracts (i.e. in this case $\text{IC}_C = \text{IC}_A$. The principal's iso-expected-profit line respective to (C) is omitted to avoid clutter.

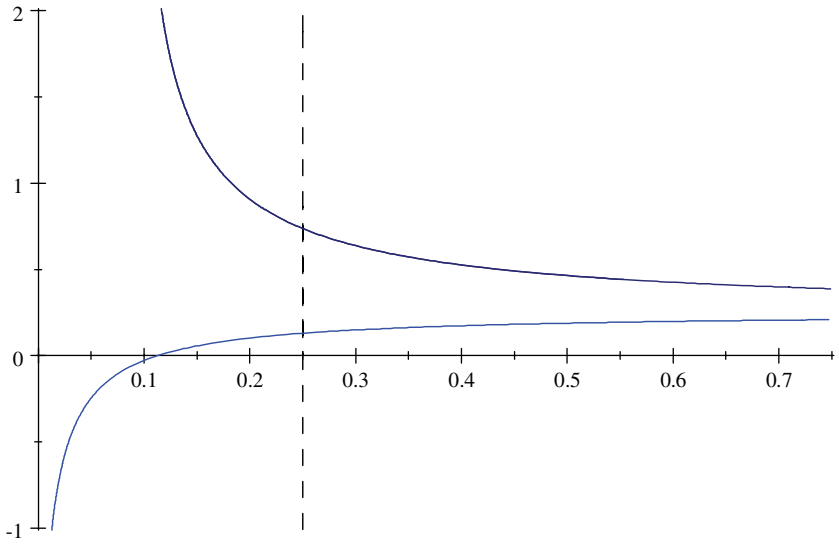

Figure 5: The *incentive effect* of overconfidence. We graph $\bar{s}_1$ (in dark blue) and $\bar{s}_0$ (in light blue) as we allow $\tilde{v}$ to vary. In this graph, we use $u(s) = 1 - e^{-s}$, $c = 0.1$, $\underline{u} = 1 - e^{-0.25}$, $q = \tilde{q} = 0.25$, and $v = 0.25$. Now $\tilde{v} = 0.25$ is the case of homogeneous beliefs, $\tilde{v} > 0.25$ means agent overconfidence, and $\tilde{v} < 0.25$ agent underconfidence. As agent overconfidence increases, a lower power of incentives is sufficient to implement effort.

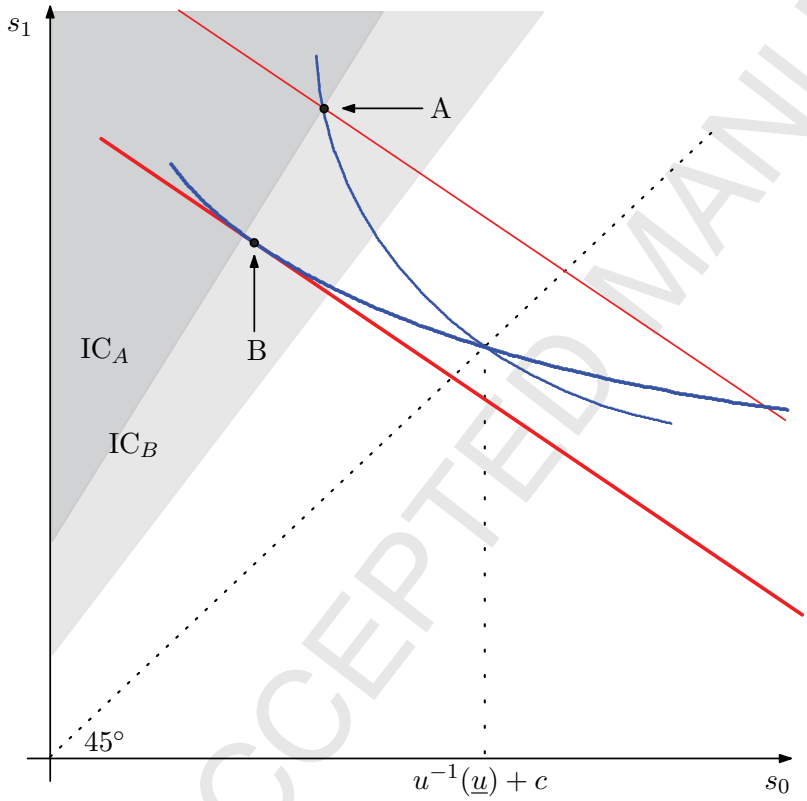

Figure 6: The second-best contract that implements effort (significant overconfidence overall). Given homogeneous beliefs, agent bears just enough risk for the contract to be incentive-compatible (A). A significantly overconfident agent bears risk according to Pareto-optimal risk sharing under disagreement (B); the power of incentives implied by such a contract is more than sufficient to implement effort, and implementation is first-best (compare to Figure 2).

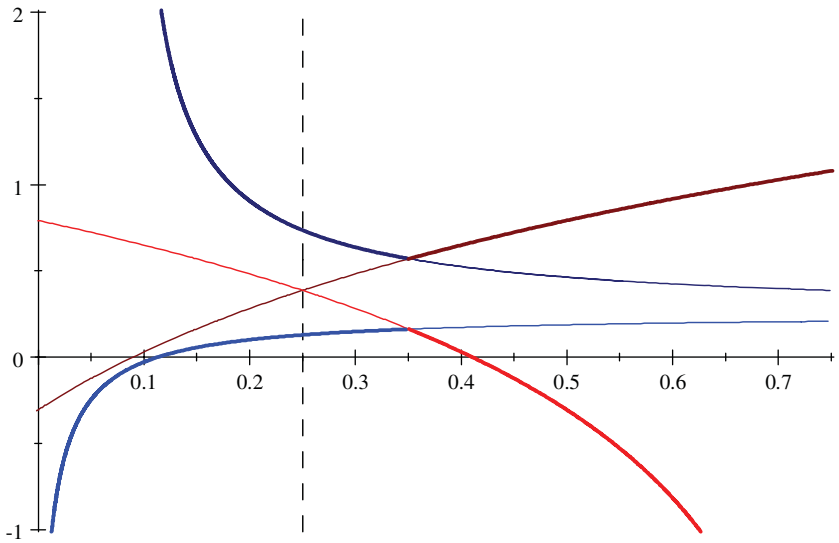

Figure 7: Overconfidence and the power of incentives. We graph s_1 (in dark blue or red depending on whether the agent is in the range of slight or significant overall overconfidence) and s_0 (in light blue or red) as we allow $\tilde{v}$ to vary. Again, we use $u(s) = 1 - e^{-s}$, $c = 0.1$, $\underline{u} = 1 - e^{-0.25}$, $q = \tilde{q} = 0.25$, and $v = 0.25$. $\tilde{v} = 0.25$ is the case of homogeneous beliefs, $\tilde{v} > 0.25$ means agent overconfidence, and $\tilde{v} < 0.25$ agent underconfidence. As overconfidence increases within the range of slight overconfidence overall, the incentive effect dominates and the power of incentives of the second-best contract decreases. As it increases further, into the range of significant overconfidence overall, the wager effect dominates and the power of incentives increases.