Overconfidence and Moral Hazard

Leonidas Enrique de la Rosa[*]
delaRosa@econ.au.dk

First version: November 2005
This version: May 16, 2008

Abstract

In this paper, I study the effects of overconfidence on incentive contracts in a moral-hazard framework in which principal and agent knowingly hold asymmetric beliefs regarding the probability of success of their enterprise. Agent overconfidence can have conflicting effects on the equilibrium contract. On the one hand, an overconfident agent disproportionately values success-contingent payments, and thus prefers higher-powered incentives. On the other hand, if the agent is overconfident in particular about the extent to which his actions affect the likelihood of success, lower-powered incentives are sufficient to induce any given effort level. If the agent is overall moderately overconfident, the latter effect dominates; because the agent bears less risk in this case, he actually benefits from his overconfidence. If the agent is significantly overconfident, the former effect dominates; the agent is then exposed to an excessive amount of risk, which is harmful to him. An increase in overconfidence—either about the base probability of success or the extent to which effort affects it—increases the effort level implemented in equilibrium.

Keywords: overconfidence, heterogeneous beliefs, moral hazard.
JEL Classification: A12, D82, D86.

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

1 Introduction

The principal-agent model has been widely used to model incentives in organizations (for a survey, see Prendergast (1999)). Some observations regarding incentive contracts remain, however, puzzling in the light of a standard principal-agent model. For instance, Prendergast (2002) notes that some empirical studies point towards a positive correlation between risk and incentives, opposite to what agency theory would predict. The heterogeneity we observe in incentive contracts (across industries and firms) does not seem to be well explained by the standard model. There is extensive psychological evidence that people are overconfident about their ability and future prospects; (De Bondt & Thaler 1995, p. 389) point out that "perhaps the most robust finding in the psychology of judgment is that people are overconfident."[1] The purpose of this paper is to study the effects of agent overconfidence in a moral-hazard model, in terms of the shape of the incentive contract and its welfare implications.

The standard moral-hazard model describes the problem facing a principal hiring an 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 profits, 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 result of this moral-hazard problem is that the owner trades 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.

This paper introduces heterogeneous beliefs of which principal and agent are aware: they "agree to disagree." The owner believes the salesman is overconfident, while he is convinced his expectations are realistic. There are two dimensions on which the asymmetry of beliefs is important in the model: an overconfident salesman can be overconfident about the *base probability of success*—the probability of success for any given effort level, which can be thought as plain optimism—and he can be overconfident about the *value of his effort*—the marginal contribution of his effort to

[1] In this paper, we will use the term "overconfidence" in reference to overestimating the probability of favorable outcomes, in particular following the agent's actions. Some literature refers to this type of self-serving bias as unrealistic optimism, positive self-image, overoptimism, or simply optimism. Others share the use of the term of overconfidence with this paper. When discussing ability and the repercussions of one's actions, I believe that overconfidence (as shorthand for excessive self-confidence) is a more appropriate term than optimism, which suggests a passive role in relation to outcomes. In the psychological literature, there is also a different use of the term overconfidence, as overestimating the precision of one's estimate.

the probability of success. Because of the parties' awareness about 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.[2] Agent overconfidence about the probability of success can have conflicting effects on the equilibrium contract. On the one hand, when the salesman is overconfident in particular about the marginal contribution of his effort to the probability of a sale, a lower commission is sufficient to induce any given effort level. I refer to this as the ***incentive effect*** of overconfidence, and it pushes the equilibrium contract to exhibit lower-powered incentives. On the other hand, because an overconfident salesman overestimates the probability of a sale, he finds 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, and it pushes the equilibrium contract to exhibit higher-powered incentives.

The degree of overall overconfidence determines which of these effects dominates in equilibrium. The incentive effect dominates when the agent is only slightly overconfident overall. When the agent is significantly overconfident, however, incentive provision becomes secondary to the fact that principal and agent value outcome-contingent payments differently. As a consequence, the wager effect dominates, and greater agent overconfidence about either the base probability of success or the effect of the agent's effort on the probability of success results in higher-powered incentives in equilibrium. Because of the potentially conflicting effects of overconfidence, the power of incentives of the equilibrium contract depends both on the degree and the kind of agent overconfidence. In contrast, the level of effort implemented by the equilibrium contract unambiguously increases with overconfidence, since both the incentive and the wager effect make higher levels of effort relatively more attractive.

Another interesting result from the model is that moderate agent overconfidence can actually be welfare-enhancing. Because an agent who is only slightly overconfident about the value of effort

[2] de Meza & Southey (1996) develop a self-selection model which results in the most overconfident entrepreneurs (which are those with the lowest probability of success) posting the highest collateral. Fang & Moscarini (2005) allow for overconfident agents who are unaware of their overconfidence, and find that a principal might prefer not to differentiate wages to avoid the negative effects that revealing her private information about the agents' true ability may have on workers' morale. Koufopoulos (2005) suggests that bias in the perception of risk might explain some empirical observations related to signaling in competitive insurance markets. Maskin & Tirole (1990) and Maskin & Tirole (1992) introduce private information held by the principal regarding the extent to which she values the agency relationship in an adverse-selection model. Villeneuve (2000) considers the possibility that the principal is better informed than the agent in an insurance-market setting, which he refers to as "reverse adverse selection."

receives more insurance than an agent who holds realistic beliefs, the gains from the reduction in the cost of agency outweigh the losses from biased estimation of his expected pay. An agent who is underconfident about the value of effort or significantly overconfident overall, in contrast, bears an excessive amount of risk, so his overconfidence will be harmful to him.

There is a growing literature that takes the possibility of overconfidence into account. Studies of overconfidence in diverse settings include Bernardo & Welch (2001) in informational cascades, Gervais & Goldstein (2007) in a model of teams with complementarities, Gervais et al. (2003) in a model of information acquisition, Goel & Thakor (2007) in tournaments, and Manove & Padilla (1999) in a screening model in which there are complementarities of effort among several agents.

This paper is closely related to contemporaneous work on overconfidence in a moral-hazard setting. Santos-Pinto (2007) 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 some conditions under which increasing overconfidence will translate into higher profits. His paper is not concerned with the shape of the incentive contract. Adrian & Westerfield (2007) 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, where the disagreement is over the project's profitability, rather than overconfidence, where the agent believes his actions are more effective than they actually are. Their results are, in the terms of this paper, about the wager effect of disagreement about the base probability of success and, importantly, of its dynamics.

The main contribution of this paper is twofold. First, it distinguishes the two relevant effects of agent overconfidence, the incentive and wager effects, in a moral-hazard setting. The distinction is important, since they are shown to have opposing consequences in terms of the power of incentives of the equilibrium contract, but they reinforce each other in terms of the implemented effort level. Second, it shows that there can be welfare gains from overconfidence in a moral-hazard setting with a risk-averse agent; these gains arise from the fundamental tradeoff between insurance and incentives, not dependent on positive externalities.

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 and several principals compete to contract with the agent. Section 4 allows for a continuum of effort levels in the agent's choice set in the alternative (dual) setting of one principal making a take-it-or-leave-it offer to the agent. Section 5 discusses the welfare effects of overconfidence. Section 6 concludes.

2 Framework

The main assumption of the model that differs 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 may be very appropriate in a moral-hazard framework; in relation to the agent's ability, arguments along the lines of "I know myself better than anybody else" for the agent and "everyone thinks they're better than average" for the principal would allow them both 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 agents' beliefs are independent of their underlying ability, she will disregard those beliefs as uninformative. In this scenario, the principal's beliefs are independent of the individual agent's true ability, so the agent can also disregard them as uninformative. Principal and agent have nothing to teach each other in terms of the agent's true ability in a one-shot game. Although this example is extreme, the important assumption is that some heterogeneity in beliefs persists after the participants 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. When discussing and agent's own ability, the psychological literature supports the case for disagreement. An alternative explanation involves errors in processing information. If players update their beliefs in a non-Bayesian way, their posterior beliefs will differ even if all private information is revealed in equilibrium. For example, the agent may overweigh success and underweigh failure when updating their beliefs about their own ability as proposed by Gervais & Odean (2001). Eyster & 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. Bénabou & Tirole (2002) model a self-deception game in which multiple equilibria regarding the level of overconfidence may arise.

In the presentation of the results, I focus on the case of *agent overconfidence*: the agent holds overly optimistic beliefs, relative to the principal, regarding the probability of success of the project.

The propositions, however, accommodate the possibility of a relatively pessimistic agent. Research in the field of psychology suggests that individuals tend to overestimate the probability of favorable events, and that such bias is more pronounced when they have some control over the 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 & Brown (1988) present a review of psychology literature that supports the view that, in general, individuals' assessment of their own abilities, talents, and social skills are overly optimistic. Support for the case of overconfidence can also be found in the business and economics literature. Larwood & 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 & Lovallo (1999), for example, find that there is excess entry into a hypothetical capacity-constrained market when participants' payoffs after entering depend on skill, but not when they depend on chance. This suggests 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."

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 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 after receiving payment s from the principal and exerting effort level e is

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

[3] This simplifying assumption ensures that the first-order approach is valid when allowing for multiple actions.

where $c(e)$ denotes the disutility to the agent from exerting effort. 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).

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 $e \in [0,1]$, 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 prove to be subsequently useful for the analysis, because it highlights the two dimensions (levels and differences) on which the asymmetry in beliefs is relevant in the model. 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 overconfident about the *base probability of success* if $\tilde{q} > q$. The agent is said to be overconfident about the *value of effort* if $\tilde{v} > v$; he overestimates the marginal contribution of his effort to the probability of success relative to the principal's beliefs. We will say that the agent is *overconfident overall* if $\tilde{q} > q$ and $\tilde{q} + \tilde{v} > q + v$. The possibility of agent underconfidence about the value of effort ($\tilde{v} < v$) is consistent with overall overconfidence and may be relevant according to some views regarding self-enhancing biases.[5] Hoorens (1993) notes that most self-enhancing biases seem to be motivated by a desire to see oneself as particularly "good" and consequently a perception of superiority (pp. 131–2). 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 underestimate the value of his effort ($\tilde{v} < v$). The agent's beliefs about the value of effort affect his perception of the rewards to 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, it is important to note the possibility of overall overconfidence coupled with underconfidence of this kind.

The solution concept used is subgame-perfect Nash equilibrium: at every decision node of the

[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.

[5] Imagine, for example, an agent who believes he has the "Midas touch": just because he's 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 underestimates the contribution of his effort to increasing the probability of success.

game, the relevant player chooses an optimal response, even if she had expected not to reach that node in equilibrium. I focus on pure-strategy equilibria of the game (in particular, each principal offers a given contract with probability one in equilibrium); this is a substantive assumption in terms of the equilibrium strategy, but does not affect the main message of the model.[6] Without loss of generality, I restrict 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.[7]

3 Binary Action Space

Consider the case of multiple principals who compete to contract with one agent. This setup is appropriate if agents are scarce. This model is also useful when considering a situation in which the agent has proprietary rights over the project, rather than the principal. Assume the agent has two actions to choose from; $e \in \{0, 1\}$.[8] A straightforward way to interpret this two-action space is that the agent can simply choose whether or not to exert effort. I normalize the cost of not exerting effort to zero so that $c(0) = 0$ and $c(1) = c$.

[6] One principal making a take-it-or-leave-it offer to the agent will always offer the optimal contract to the agent. If several principals compete to contract with the agent, equilibrium requires that the agent accepts the characterized equilibrium contract with probability 1; if the agent accepted a different contract with positive probability, there would be a profitable deviation for some principal.

[7] 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 to offer in understanding the results of this model.

[8] This assumption allows us to view the incentive and wager effects in stark contrast. Section 4 extends the model to allow for a set of (continuous) actions available to the agent. As we will see, the results and intuition extend to such a setting.

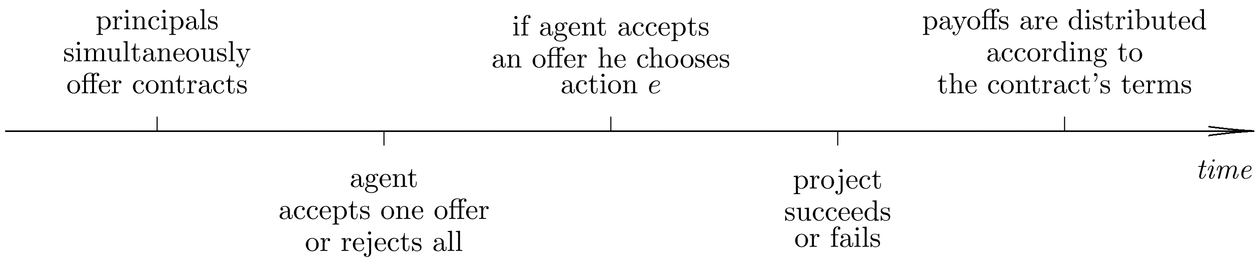

Figure 1: Timing of the model when principals compete

The timing of the model is as follows. First, principals make simultaneous contract offers to the agent. The agent then chooses which offer (if any) to accept. If the agent chooses to accept a contract offer, he chooses some action that affects the outcome distribution of the project. 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 chooses not to accept any contract, the project will not be undertaken, and the players receive payoffs according to some outside option. The outside option for each principal is not contracting with the agent, which yields zero profits. I assume that the outside option for the agent is low enough so that he always accepts an offer in equilibrium, and thus the equilibrium contract is independent of his outside option.

Principals and agent evaluate any given contract according to their own beliefs. Assume that competing principals share the same beliefs about the probability of success of the project; the possibility of different principals holding different beliefs, discussed in Subsection 5.2, does not affect the main message of the model. Each principal wishes to maximize her expected profits. Expected profits for the principal whose contract offer $\langle s_1, s_0 \rangle$ is accepted by the agent in equilibrium, conditional on each of the agent's possible effort levels, are:

$$\mathbb{E}\left[\pi \mid e=1\right] = (q+v)\left(x_1 - s_1\right) + \left[1-(q+v)\right]\left(x_0 - s_0\right)$$
$$\mathbb{E}\left[\pi \mid e=0\right] = q\left(x_1 - s_1\right) + \left[1-q\right]\left(x_0 - s_0\right).$$

The agent's objective is to maximize his expected utility when choosing which contract offer to accept and how much effort to exert once he engages the project. After accepting a given contract offer $\langle s_1, s_0 \rangle$, the agent's expected utility conditional on his choice of effort is:

$$\tilde{\mathbb{E}}\left[u\left(s_x\right) \mid e=1\right] - c = \left(\tilde{q}+\tilde{v}\right)u\left(s_1\right) + \left[1-\left(\tilde{q}+\tilde{v}\right)\right]u\left(s_0\right) - c$$
$$\tilde{\mathbb{E}}\left[u\left(s_x\right) \mid e=0\right] = \tilde{q}u\left(s_1\right) + \left[1-\tilde{q}\right]u\left(s_0\right).$$

We can now turn to characterizing the equilibrium contract. After accepting the contract offer that the agent finds most attractive, he selfishly chooses an effort level given the terms of the contract. The competing principals take this into account when designing their offers. In particular, if a principal wishes to induce effort, the contract must be incentive compatible—the contract terms must be such that the agent finds it in his best interest to exert effort:

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

We can rewrite the incentive-compatibility constraint above as

$$\tilde{v}\left(u\left(s_1\right)-u\left(s_0\right)\right)\geq c. \tag{1}$$

Intuitively, the perceived expected utility gain for the agent from exerting effort (receiving excess utility $\left(u\left(s_1\right)-u\left(s_0\right)\right)$ with additional probability $\tilde{v}$), must be no less than his disutility from exerting effort. I will refer to the differential $u\left(s_1\right)-u\left(s_0\right)$ as the contract's *power of incentives*. Note that the power of incentives necessary to induce effort is decreasing in $\tilde{v}$. It is, however, independent of $\tilde{q}$: the agent's beliefs regarding the base probability of success does not affect his perception of the rewards to effort of a given incentive scheme.

If the principal does not wish to induce effort, her only concern will be to attract the agent into accepting her contract offer.

An equilibrium contract is such that no other contract can *(i)* attract the agent by offering him terms that he strictly prefers and *(ii)* yield higher expected profits for the offering principal (i.e. there is no profitable deviation for any principal from an equilibrium contract). Because principals compete in offering contracts to the agent and they all evaluate profits based on the same beliefs, expected profits for the principal whose offer is accepted by the agent in equilibrium must be zero (equal to the principals' outside option).

Lemma 1 *If principals share the same beliefs regarding outcome distribution conditional on the agent's actions, in equilibrium expected profits will be zero for all principals according to their beliefs.*

All proofs are relegated to the appendix. Intuitively, if a principal made positive expected profits, another principal could outbid that contract offer—provide a slightly higher expected payment to the agent—while implementing the same action, thus attracting the agent and earning positive expected profits.

For exposition, I will in turn characterize the equilibrium contract assuming that effort is not implemented and assuming that effort is implemented, and subsequently analyze the effect of overconfidence on the level of effort actually implemented in equilibrium.

Assume first that effort is not implemented in equilibrium. If principals and agent held identical beliefs, the risk-neutral principal would absorb all the risk from the project, and offer a fixed payment to the agent (i.e. independent of project outcome). If the agent is overconfident about the base probability of success, however, he will be exposed to risk in equilibrium.

Proposition 1 *Assuming effort is not implemented in equilibrium, the only equilibrium contract* $\langle s_{1*}, s_{0*} \rangle$ *is characterized by the conditions*

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

and $q(x_1 - s_{1*}) + [1-q](x_0 - s_{0*}) = 0.$

Consider the case of agent overconfidence about the base probability of success, in which $\tilde{q} > q$. Proposition 1 is simply an application of the Borch (1962) rule for optimal risk-sharing, allowing for heterogeneous beliefs. The intuition of is that an overconfident agent is willing to wager on success against the (relatively pessimistic) principal. Starting from a riskless contract (one that specifies $s_1 = s_0$), because the marginal cost for the agent from bearing additional risk is zero at that point, principal and agent evaluate marginal changes in payments based on their effect only in terms of expected payment. 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 principals' beliefs. The agent is relatively optimistic about receiving the success-contingent payment, so *according to his beliefs* such deviation yields a higher expected payment. When $\tilde{q} > q$, there is a first-order gain perceived by the agent from such higher expected payment, and only a second-order loss from higher risk exposure when evaluated at the riskless contract. Therefore, an agent who is overconfident about the base probability of success bears risk in equilibrium.

Because of the disagreement between principal and agent regarding the probability of success of the project, the agent is willing to be exposed to more risk in equilibrium than a "realistic" agent. This ***wager effect*** pushes the equilibrium contract towards higher-powered incentives. Note that, 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.[9]

Assume now that effort is implemented in equilibrium. If principals and agent held identical beliefs, the equilibrium contract offer would be characterized by zero expected profits for the of-

[9] If the contract characterized in Proposition 1 satisfies the incentive-compatibility constraint (1), it must be the case that the agent exerts effort in equilibrium. It will become clear that this is the extreme case of the positive effect of overconfidence about the base probability of success on the implemented level of effort (see Proposition 4).

fering principal and the binding incentive-compatibility constraint (1). There is a tradeoff between incentives and insurance: if not for the incentive-provision problem, the principal would like to offer more insurance to the agent (i.e. reducing the power of incentives). Implementing effort requires that the agent be exposed to a discrete amount of risk. 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 (for instance, 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 equilibrium 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* (1) *with equality and yields zero expected profits according to the principals' beliefs:*

$$\tilde{v}\left(u\left(\bar{s}_1\right) - u\left(\bar{s}_0\right)\right) = c$$
$$(q+v)\left(x_1 - \bar{s}_1\right) + \left[1 - (q+v)\right]\left(x_0 - \bar{s}_0\right) = 0.$$

This contract will in fact be the equilibrium contract when the beliefs held by the agent differ only slightly from the principals' beliefs. In other words, if the agent is only slightly overconfident overall, the contract with incentives just powerful enough to implement effort will be the equilibrium contract. The intuition from the identical-beliefs setting carries over to this case: a principal cannot provide more insurance to the agent without destroying the incentives for the agent to exert effort.

Proposition 2 *Assuming effort is implemented in equilibrium, if* $\frac{\tilde{q}+\tilde{v}}{1-(\tilde{q}+\tilde{v})}\frac{u'(\bar{s}_1)}{u'(\bar{s}_0)} \leq \frac{q+v}{1-(q+v)}$ *then* $\langle \bar{s}_1, \bar{s}_0 \rangle$ *is the only equilibrium contract.*

Given that $\bar{s}_1 > \bar{s}_0$, the condition $\frac{\tilde{q}+\tilde{v}}{1-(\tilde{q}+\tilde{v})}\frac{u'(\bar{s}_1)}{u'(\bar{s}_0)} \leq \frac{q+v}{1-(q+v)}$ holds for some values $\tilde{q}+\tilde{v} > q+v$. If this is the case, we will say that the agent is only "slightly overconfident overall," since $\tilde{q}+\tilde{v} \ggg q+v$. There is no profitable deviation from $\langle \bar{s}_1, \bar{s}_0 \rangle$. Providing more insurance to the agent (by decreasing s_1 and increasing s_0) would destroy the incentives for the agent to exert effort, while it would be too costly for the principal to compensate the agent for bearing more risk. Given the condition above, the agent is already bearing too much risk compared to what would be optimal risk-sharing.

When the agent is only slightly overconfident overall, his overconfidence about the value of effort ($\tilde{v} > v$) allows a principal to provide him more insurance without destroying incentives compared to a "realistic" agent. This is the ***incentive effect*** of overconfidence. 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 is willing to bear is continuous in the degree of disagreement in beliefs. 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 willingly bear because of disagreement. The incentive effect therefore dominates the wager effect in this case, and the power of incentives of the equilibrium contract depends solely on the agent's beliefs about the value of effort.

When the agent is only slightly overconfident overall, the incentive-insurance tradeoff present in the case of identical beliefs remains. If the agent is overconfident about the value of effort, the principal can provide more insurance (thus reducing the cost of agency) without destroying incentives. If, however, the agent is underconfident about the value of effort, the incentive effect implies that higher-powered incentives will be necessary to implement effort. Applying Lemma 1, we can see that $\frac{d\bar{s}_1}{d\tilde{v}} < 0$ and $\frac{d\bar{s}_0}{d\tilde{v}} > 0$ when the agent is only slightly overconfident overall. Therefore, **the power of incentives of the equilibrium contract decreases in overconfidence about the value of effort when the agent is only slightly overconfident overall**. The agent's beliefs regarding the base probability of success ($\tilde{q}$) do not affect the equilibrium contract in this case. The principals' beliefs do not affect the power of incentives of the equilibrium contract either, but they do affect the agent's expected payment.

The degree of asymmetry in beliefs held by principals and agent defines whether or not the agent is only slightly overconfident overall. As the next proposition shows, if the agent is instead significantly overconfident overall, the equilibrium contract exhibits excessively powerful incentives. Because of the wager effect, a very overconfident agent substantially overestimates the probability of success, so 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). He judges the higher expected payment from an excessively risky contract as sufficient to compensate him for the cost of bearing more risk.

Proposition 3 *Assuming effort is implemented in equilibrium, if* $\frac{\tilde{q}+\tilde{v}}{1-(\tilde{q}+\tilde{v})}\frac{u'(\bar{s}_1)}{u'(\bar{s}_0)} > \frac{q+v}{1-(q+v)}$ *then* $\langle \bar{s}_1, \bar{s}_0 \rangle$ *is not an equilibrium contract. The only equilibrium contract* $\langle s_1^*, s_0^* \rangle$ *is characterized by*

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

and $(q+v)(x_1 - s_1^*) + [1-(q+v)](x_0 - s_0^*) = 0$. *The equilibrium contract has higher-powered incentives than necessary to implement effort.*

If $\tilde{q}+\tilde{v} \gg q+v$, so that $\frac{\tilde{q}+\tilde{v}}{1-(\tilde{q}+\tilde{v})}\frac{u'(\bar{s}_1)}{u'(\bar{s}_0)} > \frac{q+v}{1-(q+v)}$, we will say that the agent is "significantly overconfident overall." If this is the case, he is actually content to bear more risk in equilibrium than he would under 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 more than compensates him for the excessive amount of risk he bears.

When the agent is significantly overconfident overall, the agent's bias in evaluating payments overshadows the incentive-insurance tradeoff present in the identical-beliefs case. Incentive provision becomes secondary to the difference in how principal and agent evaluate outcome-contingent payments. Optimal risk-sharing leads to the agent bearing more risk than necessary to implement effort.

In contrast to the effect of overconfidence about the value of effort on the power of incentives when the agent is slightly overconfident, now $\frac{ds_1^*}{d\tilde{v}} > 0$ and $\frac{ds_0^*}{d\tilde{v}} < 0$. Likewise, $\frac{ds_1^*}{d\tilde{q}} > 0$ and $\frac{ds_0^*}{d\tilde{q}} < 0$. Because it is the wager effect of overconfidence that dominates in this case, **the power of incentives of the equilibrium contract increases in overconfidence of either kind when the agent is significantly overconfident overall**.

Whether or not effort is implemented in equilibrium depends on which of the potential equilibrium contracts identified above gives the agent higher perceived expected utility (recall that principals always receive zero expected profits in equilibrium). The incentive effect reduces the cost of implementing effort, and a higher degree of overconfidence of either kind makes implementing effort relatively more attractive to the agent. Therefore, **higher levels of overconfidence of either kind increase the likelihood that effort is implemented in equilibrium**.

Proposition 4 *Ceteris paribus, higher levels of overconfidence of either kind increase the likelihood that effort is implemented in equilibrium: 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* $\tilde{q}^* \geq \tilde{q}$.

When the incentive effect dominates, higher overconfidence about the value of effort reduces the cost of agency. Higher overconfidence about the base probability of success increases the agent's perceived expected utility under the contract that *does not* implement effort; however, because he receives a higher success-contingent payment under the contract that *does* implement effort, there is a greater increase in his perceived expected utility under this contract. For this same reason, higher overconfidence of either kind makes the contract that implements effort comparatively more attractive to the agent when the wager effect dominates.[10]

[10] Note that Proposition 4 does not imply that effort is more likely to be implemented, in general, when dealing with an overall more overconfident agent. Imagine, for example, a case in which effort would be implemented if the agent held "realistic" beliefs. It is here that the possibility of underconfidence about the value of effort coupled with overall overconfidence is relevant. Because higher-powered incentives are then necessary to implement effort, the increase in the cost of agency might make effort too costly to implement.

4 Continuous Action Space

Consider now the setting that tends to be discussed more often in agency literature, in which one principal can make a take-it-or-leave-it contract offer to one agent. Owing to her bargaining power, the principal extracts all the surplus from the agency relationship. This framework is more appropriate than the competing-principals setting when the pool of potential agents is large relative to the number of principals. Assume that the agent can choose some $e \in [0,1]$ if he accepts the principal's contract offer. The disutility cost of effort is a function $c(e)$; assume that $c'(\cdot) > 0$ and $c''(\cdot) > 0$. This assumption 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 $s_1 > s_0$.[11]

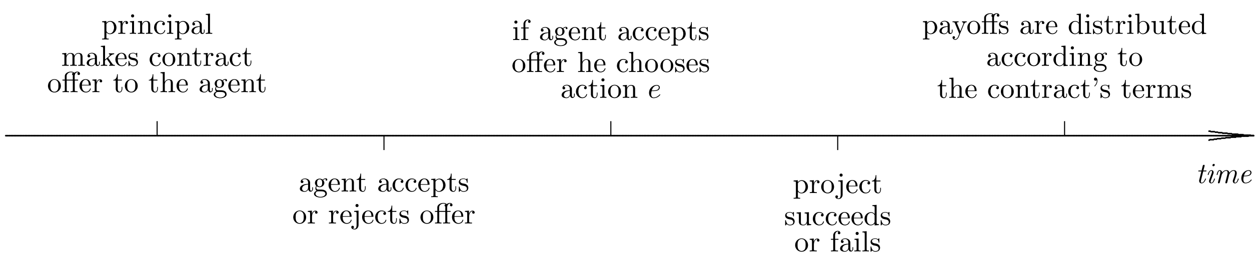

Figure 2: Timing of the model with one principal and one agent

The timing of the model is as follows. 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 then realized, payoffs are distributed according to the terms of the contract, and the agency relationship ends. If the agent rejects the principal's offer, both will receive utility according to their outside option. I assume that the agent's outside option is exogenous and independent of his overconfidence.[12] The principal's outside option does not affect the equilibrium contract as long as the agency relationship yields sufficient surplus for her to engage in it; I assume this to be the case.

Let $\underline{u}$ denote the agent's perceived expected utility from his outside option. The principal's

[11] Note that if the agent chose 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 two-action model.

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

contract offer, if it is to be accepted by the agent, must provide him 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

$$(\tilde{q} + \tilde{v}e)\, u\,(s_1) + [1 - (\tilde{q} + \tilde{v}e)]\, u\,(s_0) - c\,(e) \geq \underline{u}, \tag{2}$$

where e is the action (freely chosen by the agent) that the principal wishes to implement. The principal who can make a take-it-or-leave-it offer to an agent faces a very similar problem to the one faced by a principal who competes with others. Intuitively, there is no profitable deviation from the equilibrium contract in the competing-principals setting—there is no other contract that attracts the agent (gives him higher perceived expected utility) and yields higher expected profits for the offering principal. The optimal contract when one principal makes a take-it-or-leave-it offer is such that no other contract retains the agent (gives him at least as much perceived expected utility as his outside option) and yields higher expected profits for the principal.

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:

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

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

$$\tilde{v}\,[u\,(s_1) - u\,(s_0)] = c'\,(e)\,, \tag{3}$$

which defines the agent's choice of effort after accepting contract offer $\langle s_1, s_0 \rangle$.[13] 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.

Note that the agent's participation constraint (2) must be binding in equilibrium. If not, the principal could marginally reduce both payments s_1 and s_0 while keeping the power of incentives $[u\,(s_1) - u\,(s_0)]$ constant (so as to implement the same effort level), increasing her expected profits.

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 (2) and the incentive-compatibility constraint (3) characterize the least-cost contract for the principal to implement any given effort level e.

Definition 2 *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*,* $\langle s_1\,(e)\,, s_0\,(e) \rangle$*, is*

[13] Simplifying the outcome space to success and failure ensures that the first-order approach is valid.

implicitly defined by

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

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 $s_1(e)$ and $s_0(e)$ are implicitly defined above. Clearly, the power of incentives of the optimal contract depends not only on the agent's beliefs but also on 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, it is possible to study the qualitative effects of changes in each kind of overconfidence.

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:

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

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 depends crucially on the agent's beliefs. It is clear that for the principal's profit-maximization problem to be well behaved, it is sufficient that the marginal cost of implementing effort be an increasing function of effort level. Further discussion about the conditions under which the principal's profit-maximization problem is well behaved is relegated to the Appendix (Subsection A.2.1).

Consider the effect of marginally higher agent overconfidence about the base probability of success on the marginal cost of implementing effort, evaluated at the optimal e^*:

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

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 an agent who is more overconfident about the base probability of success**: $\frac{de^*}{d\tilde{q}} > 0$. This result is analogous to its counterpart in the two-action case, summarized in Proposition 4. As a consequence of the wager effect of overconfidence, because a more-overconfident 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 marginally higher agent overconfidence about the value of effort:

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

The first term of the equation above reflects the wager effect of overconfidence. Just as in the case of overconfidence about the base probability of success, 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 benefits the principal as long as $(x_1 - s_1(e^*)) \geq (x_0 - s_0(e^*))$.[14] 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, **higher overconfidence about the value of effort results in a higher**

[14] It seems reasonable to assume that in equilibrium $x_1 - x_0 \geq s_1(e^*) - s_0(e^*)$, so that the agent is not exposed to more risk than inherent in the project. If this assumption is violated, the principal would enjoy greater profits in the case of project failure than in the case of success, so may prefer to implement lower rather than higher effort levels. Such a contract seems unrealistic, since the principal would then have incentives to sabotage the project. Common sense, if not legal restrictions, should prevent the agent from accepting such a contract.

implemented effort level: $\frac{de^*}{d\tilde{v}} > 0$. This is analogous to the corresponding result in the two-action case, exposed in Proposition 4.

The comparative statics of the power of incentives of the optimal contract with respect to overconfidence about the base probability of success is straightforward. Given that a higher effort level is implemented, and that this kind of overconfidence does not directly affect the incentive structure of the contract holding effort constant, it follows that overconfidence about the base probability of success always implies higher-powered incentives. The wager effect of overconfidence drives the optimal contract towards higher-powered incentives in the continuous-action case, even if the agent is only slightly overconfident overall, because the implemented effort level continuously increases with overconfidence.

The comparative statics of the power of incentives of the optimal contract with respect to overconfidence about the value of effort is more subtle, and formally discussed in the Appendix (Subsection A.2.2). The incentive effect of overconfidence still pushes towards lower-powered incentives, but both the wager effect and the increase in the implemented level of effort push towards higher-powered incentives. There will be some range of slight overconfidence over which the power of incentives decreases with overconfidence about the value of effort if the agent is sufficiently risk averse and the marginal cost of exerting effort does not increase sharply. Intuitively, because a small increase in the amount of risk born by the agent is sufficient to induce higher effort exertion, and lower-powered incentives are sufficient to implement any given effort level, the power of incentives of the optimal contract decreases with overconfidence about the value of effort. If, on the other hand, the agent is not very risk averse or his marginal cost of exerting effort increases sharply, the additional power of incentives necessary to implement higher effort is large, so there is no such range and the power of incentives is everywhere increasing in overconfidence of either kind.

4.1 Choosing From a Pool of Agents

Given that agent beliefs affect the principal's expected profit—even if we hold actual agent productivity constant—the question regarding whether the principal prefers to contract with a more- or less-overconfident agent follows naturally. Consider the problem that the principal faces when choosing an agent from a pool of applicants with different levels of overconfidence and true ability. This question is particularly relevant when the pool of agents is large relative to the number of principals willing to hire an agent, our main motivation for discussing the one-principal setting.

All else equal, the principal will prefer a more overconfident agent. Imagine a situation in which many potential agents share the same underlying characteristics, but differ in their self-confidence—e.g. a group of individuals who pass several aptitude tests. The principal will choose the most overconfident agent, because it is cheapest for her to satisfy his participation constraint.[15] Clearly, the assumption that the outside option is evaluated equally by agents with different beliefs is crucial for this result: less overconfident agents could become attractive if overconfidence also affects their perception regarding their outside option, since less overconfident agents would accept contracts yielding lower perceived expected utility.

When facing a pool of agents who share the same beliefs about their ability, on the other hand, the principal will hire the agent that she judges to be most able. An applicant who responds to a job announcement, for example, probably believes he is well-suited for the position. Because choosing a higher-ability agent (i.e. an agent who generates a better outcome distribution) yields higher expected revenue and the cost of inducing any of these agents to exert effort is the same, the principal naturally prefers the most able agent.

In short: When the principal faces a pool of same-ability agents who differ only in their level of overconfidence, she will tend to hire the *most* overconfident agent. In contrast, when she faces a pool of applicants who share the same beliefs but differ in underlying ability, the principal will choose the *least* overconfident—most able—agent.

5 Welfare Analysis

An overconfident agent will have a biased outlook on his expected utility; in what follows, I evaluate the agent's *actual expected utility* based on the true probabilities of success and failure. Although there are strong arguments for using subjective expected utility, side-betting among disagreeing individuals always allows for *ex-ante* Pareto gains. I will focus instead on the potential *ex-post* Pareto gains. Further considerations about factors that influence individuals' well-being have conflicting implications in terms of welfare analysis. If the agent derives utility from anticipating how richly he will be rewarded once the project succeeds, along the lines of Kőszegi (2005), an overconfident agent will enjoy higher utility than what I calculate as his actual expected utility. On the other hand, once payoffs are realized, the agent will be disappointed whenever he does not receive the success-contingent payment he so confidently anticipates. An agent could evaluate receiving the low failure-contingent payment not only for its own worth, but also as a loss relative to his "reference point" as introduced by Kőszegi & Rabin (2006). In this case, the agent would be worse off than what I calculate.

[15] As discussed before, if an agent is overconfident overall but underconfident about the value of effort, the cost of agency increases. The principal might choose a less (overall) overconfident agent if lower-powered incentives are sufficient to induce this agent to exert effort. More precisely, fixing actual agent ability and one kind of overconfidence across agents, the principal will prefer the agent with highest overall overconfidence.

As will become clear, any possible (*ex-post*) gains must come about when the incentive effect of overconfidence dominates and the cost of agency decreases with respect to the rational benchmark. The fact that lower power of incentives is sufficient to implement any given level of effort reduces the inefficiency inherent in the agency relationship when a risk-averse agent bears risk in equilibrium. If the principal receives all the gains from trade, she will benefit from any efficiency gains,[16] and the agent can only lose from biased beliefs. If, however, the agent receives the gains from trade, then he will capture the gains from a lower cost of agency. The binary-action setting studied in Section 3 allows for a straightforward exposition of this fact.

5.1 Principals Sharing the Same Beliefs

Assume, for now, that the principals hold accurate beliefs. This assumption is convenient because the zero-expected-profits condition implies that social welfare depends solely on the agent's well-being and that the agent's beliefs only affect the power of incentives of the equilibrium contract—not the actual expected payment the agent receives. We will relax this assumption in the following Subsection, noting the complications that arise.

When principals compete in making offers to an agent or the agent makes a take-it-or-leave-it offer to a principal, the agent receives expected payment equal to the project's expected revenue. Because, given some implemented level of effort, the agent's expected payment is independent of the terms of the equilibrium contract, his actual expected utility depends exclusively on the implemented level of effort and the amount of risk he bears in equilibrium.

Keeping with the structure of Section 3, I first identify the welfare effects of overconfidence assuming that effort is not implemented in equilibrium, and that changes in overconfidence do not affect the implemented level of effort. In that case, zero expected profits and the assumption that principals hold accurate beliefs imply that actual expected payment to the agent is the project's expected revenue:

$$qx_1 + (1-q)\,x_0.$$

As shown in Proposition 1, as a consequence of the wager effect, an agent who is overconfident about the base probability of success always bears risk in equilibrium, so any level of overconfidence of this kind harms the risk-averse agent.

Assume now that effort is implemented in equilibrium, and that changes in overconfidence do not affect the implemented level of effort. Actual expected payment to the agent is then

$$(q+v)\,x_1 + [1-(q+v)]\,x_0.$$

[16] This is the main result in Santos-Pinto (2007): he derives conditions sufficient for the principal's profits to be increasing in overconfidence.

When the agent is only slightly overconfident overall, the equilibrium contract is characterized by Proposition 2. If this is the case, the power of incentives depends solely on the agent's beliefs regarding the value of effort. Due to the incentive effect, an agent who is overconfident about the value of effort is exposed to less risk in equilibrium than he would be if he held realistic beliefs because lower-powered incentives are sufficient to induce effort. Thus, the agent's well-being *increases* with overconfidence regarding the value of effort. Slight overconfidence about the value of effort is therefore beneficial to the agent: he benefits from the efficiency gain of a lower cost of agency. The agent's beliefs about the base probability of success do not affect the equilibrium contract if he remains only slightly overconfident overall.

When the agent is significantly overconfident overall, the equilibrium contract is characterized by Proposition 3. As a consequence of the wager effect, the amount of risk that the agent is exposed to increases in overconfidence of either kind. Thus, when the agent is significantly overconfident, his well-being *decreases* with overall overconfidence. Higher overconfidence increases the agent's exposure to what is already an excessive amount of risk.

As shown by Proposition 4, another effect of higher overconfidence of either kind is that it makes effort more likely to be implemented. If higher agent overconfidence drives effort to be implemented in equilibrium, the actual expected payment to the agent increases (by as much as expected revenue does) because effort exertion increases the probability of success of the project. A marginal increase in overconfidence that drives effort to be implemented might, therefore, benefit the agent. This will only be the case, however, if the increase in actual expected payment compensates the agent for the disutility of exerting effort (effort must therefore be the first-best action, so that $v(x_1 - x_0) > c$) and the additional risk he bears.

When principals compete to contract with the agent, some overconfidence about the value of effort benefits the agent because lower-powered incentives are then sufficient to induce effort. A risk-neutral principal can provide more insurance without destroying incentives.[17] In contrast, if the agent is underconfident about the value of effort or significantly overconfident overall, he bears an excessive amount of risk (which is costly to him).

Allowing for principals to hold inaccurate beliefs, the effects of overconfidence outlined above are reinforced if principals are overly optimistic about the probability of success of the project. When this is the case, higher-powered incentive contacts that yield zero expected profits according to the principals' beliefs yield lower actual expected payment to the agent. As the agent holds

[17] Note that when we allow for a continous action space, the conditions that ensure a decreasing power of incentives as a function of overconfidence about the value of effort discussed in the Appendix (Subsection A.2.2) are sufficient to guarantee *ex-post* Pareto gains from moderate levels of overconfidence.

higher stakes in the project, his actual expected payment decreases from the optimistic principals' estimate of expected revenue towards actual expected revenue.[18] The agent's welfare therefore decreases as the power of incentives of the equilibrium contract increases, even more sharply than in the case in which principals hold accurate beliefs. The effects of agent overconfidence on his well-being are therefore reinforced when principals are overly optimistic: the agent benefits from moderate overconfidence about the value of effort because he bears less risk *and* he receives a higher actual expected payment in equilibrium. Underconfidence about the value of effort or significant overall overconfidence harm the agent both because he is exposed to a higher amount of risk and because he receives a lower actual expected payment in equilibrium.

If the principals are overly pessimistic relative to the true probability of success, the effects of agent overconfidence on his well-being are ambiguous. In this case, when expected profits are zero according to the principals' beliefs, higher-powered incentive contracts yield higher actual expected payment to the agent. As the agent holds higher stakes in the project, his actual expected payment now increases from the overly pessimistic principals' estimate of expected revenue towards actual expected revenue. Changes in overconfidence that result in lower power of incentives still benefit the agent by shielding him from risk, but provide him with a lower actual expected payment. Conversely, changes in overconfidence that result in higher power of incentives expose the agent to more risk but also provide higher actual expected payment to him. For instance, a significantly overconfident agent may actually benefit from his overconfidence when contracting with an overly pessimistic principal, particularly if his beliefs are close to the true outcome distribution and he has high risk tolerance.

5.2 Principals With Differing Beliefs

I derived the results of the model assuming that all competing principals hold identical beliefs. This simplified the analysis because competition then resembles Bertrand competition, so that expected profits are driven to zero when as few as two principals compete. Given that I am allowing for principal and agent to hold heterogeneous beliefs, this assumption seems particularly strong. If we allow for principals to knowingly hold disagreeing beliefs regarding the probability distribution of outcomes, two main complications arise.[19]

[18] This is true as long as $s_1 < x_1$ and $s_0 > x_0$. I assume that this is the case, but briefly discuss the alternative in Subsection 5.2 below.

[19] When the principals hold heterogeneous beliefs, it would be optimal for them to set up a secondary side-betting market on project outcome (if it is publicly observable). Because of risk neutrality, this side-betting would be unbounded—and so would expected profits from each principal's point of view. We assume that such a side-betting market is infeasible, so that principals behave optimally in the contract-offer-design stage.

First, so long as the number of competing principals is finite, expected profits according to the beliefs of the principal whose offer is accepted in equilibrium will generically be positive. The most optimistic principal needs only to ensure that her contract offer yields zero expected profits according to the beliefs of the second-most-optimistic principal. The second-most-optimistic principal is not willing to outbid some offers that yield positive expected profits to the most optimistic principal if they hold heterogeneous beliefs.[20] When allowing for principals to hold disagreeing beliefs, it is generally in each principal's best interest to hold accurate beliefs, which will allow her to design her contract offer optimally. If the second-most-optimistic principal overestimates the true probability of success, the most optimistic principal (whose contract is accepted in equilibrium) will suffer losses in expectation. Optimistic bias will therefore tend to harm the principal. An overly pessimistic principal, on the other hand, will tend to be outbid by more optimistic principals. If all principals are pessimistic relative to the actual probability of success, any principal whose offer is rejected in equilibrium would benefit from correctly updating her beliefs; she could then attract the agent and earn positive expected profits.

The second complication is pinpointing the offer which is accepted by the agent in equilibrium. If we assume monotonic bidding strategies by the principals, then the offer made by the most optimistic principal will tend to be the one accepted in equilibrium. There is, however, an extreme case in which the offer made by the most *pessimistic* principal could be the one accepted by the agent: if the agent is fairly risk neutral and very overconfident relative to the most pessimistic principal, the $\langle s_1^*, s_0^* \rangle$ contract as characterized in Proposition 3 could be such that $s_1^* > x_1$ and $s_0^* < x_0$—the principal bets on project failure.[21] If this is the case, the principal earns higher profits when the project fails than when it succeeds. The principal would then have incentives to sabotage the project if possible, and the agent should be wary of accepting such a contract offer. This extreme seems unrealistic, for the same reason that contracts with payments that are non-monotonic in the principal's objective variable (e.g. output) seem unrealistic: one of the participants would then have incentives to destroy output.

[20]Note that the result of generically-positive perceived expected profits for the principal whose offer is accepted in equilibrium does not depend on the assumption that principals are aware of each others' beliefs. Dropping this assumption (if principals know only the agent's and their own beliefs), competition would then resemble a first-price sealed-bid auction. If there are finitely many principals, each of them knows that some contract offers which yield strictly positive expected profits will be accepted by the agent with positive probability. Each principal will therefore offer a contract to the agent that, conditional on being accepted, gives the principal strictly positive perceived expected profits according to her own beliefs.

[21]This case may be so extreme that holding excessively pessimistic beliefs could be better than holding accurate beliefs for a principal: she then "wins" and earns the chance to bilk the agent for profit, whereas the offer made by a principal holding accurate beliefs would be rejected.

6 Conclusion

This paper attempts to provide insight into the effects of overconfidence in equilibrium within a moral-hazard framework. It gives one possible explanation to why we may observe incentive contracts that seem excessively powerful in some situations, and others that seem surprisingly flat in other situations. It also shows that, in an agency setting, overconfidence can be valuable for the agent when the principal is aware of such bias.

The results of the paper suggest that incentive contracts are sensitive to the *kind* 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 & 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 about the value of effort, less powerful incentives would follow.[22]

Because of the non-monotonic relationship between overconfidence and power of incentives, the results of this paper give a rationale for some of the contradictory empirical evidence relating power of incentives to uncertainty. (Prendergast 2002, p. 1073) notes that "the evidence on executive compensation about whether risk and incentives are substitutes is mixed. Some studies find evidence in favor of the negative trade-off, others find evidence of a positive relationship, and some find no relationship." Following (Morris 1995, p. 238), we can argue that in "risky" situations, where learning has ceased, we should expect beliefs to have converged, and thus to see the negative correlation between risk and incentives predicted by the standard model. In "uncertain" situations, where learning has not ceased and there is scope for disagreement, our model predicts that there will be room for the seemingly inconsistent results. Given that we expect to see more heterogeneity of beliefs in more uncertain environments, the sign of the correlation between the power of incentives and uncertainty depends on whether the incentive or the wager effect dominates in a given environment. Along these lines, we can also reinterpret one of the observations made

[22] The type of 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 & Dickens (1982). My model implies, however, that different kinds of overconfidence can have conflicting effects in terms of the amount of risk born by the agent in equilibrium. If this is the case, agents with similar degrees of overall overconfidence might sort themselves into very different positions.

by Smith & Watts (1992) regarding executive compensation. They note that firms with larger investment opportunity sets pay their CEOs more, and are more likely to use stock options and other forms of performance-contingent pay. We should expect that, as argued by Van den Steen (2004), larger investment opportunity sets are coupled with larger CEO overconfidence. If the wager effect dominates, we should in fact expect to see contracts with more powerful incentives in industries where the firms face larger investment opportunity sets.[23]

We may also ask whether overconfidence is related to the probability of success of agents. When testing for such a relationship in the field, the unobserved heterogeneity across agents both in terms of ability and beliefs may lead to identification problems. In their survey of almost three thousand entrepreneurs, Cooper et al. (1988) uncover a regularity that may help overcome this problem. They find that the entrepreneurs' degree of overconfidence seems to be independent of factors that affect their actual probability of success (like experience in the industry and education level). This suggests an empirical test for the relevance of overconfidence in entrepreneurs' decision making: holding the beliefs of the entrepreneur constant, if the most overconfident agents are *significantly overconfident* then we may find that the probability of success of an enterprise may be *negatively* correlated to the proportion of the entrepreneur's own funds invested in it, similar to what de Meza & Southey (1996) suggest would happen in a self-selection model with overconfident agents.

The model presented here does predict a monotonic relationship between overconfidence and the level of effort implemented in equilibrium. 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 explanation of Weinstein (1980), applied to this case, would be as follows: when entrepreneurs form expectations they may be comparing themselves to a hypothetical entrepreneur who chooses to enter an industry with merely good prospects and puts little effort into making the enterprise succeed. Given that the average success rate of businesses is readily available information, entrepreneurs may use this average as their benchmark when forming expectations. When doing so, they fail to internalize the fact that most other entrepreneurs choose to enter an industry which they deem particularly profitable and work hard towards success, just like they do. The results of my model imply that entrepreneurs' choice of long hours may be well explained by their overconfidence.

[23] Note that this is somewhat at odds with the results of Gervais et al. (2003), who argue for smaller option packages when dealing with overconfident managers. Their argument is that lower-powered incentives are sufficient to align the agent's objectives to those of the principal (the incentive effect). When we take the wager effect into account, however, we might expect to see higher-powered incentive contracts in equilibrium.

A Appendix

A.1 Binary Action Space

Lemma 1 *If principals share the same beliefs regarding outcome distribution conditional on the agent's actions, in equilibrium expected profits will be zero for all principals according to their beliefs.*

Proof. Suppose not. Suppose that a principal offers, and the agent accepts, a contract $\langle s_1, s_0 \rangle$ in equilibrium that yields positive expected profits for the offering principal:

$$\Pr(x_1 \mid e)(x_1 - s_1) + [1 - \Pr(x_1 \mid e)](x_0 - s_0) > 0,$$

where e is chosen optimally by the agent. Other principals, whose contracts are not accepted by the agent, receive zero profits in equilibrium. Consider the following deviation by one of these principals: offer contract $\langle s_1 + \epsilon, s_0 \rangle$ if $\langle s_1, s_0 \rangle$ implements effort, and $\langle s_1, s_0 + \epsilon \rangle$ if $\langle s_1, s_0 \rangle$ does not implement effort. In either case, the new contract implements the same effort level as $\langle s_1, s_0 \rangle$ does, makes the agent strictly better off, and will thus be accepted by the agent. For ϵ close enough to zero, the principal making the new contract offer enjoys positive expected profits. Given that there is a profitable deviation, $\langle s_1, s_0 \rangle$ cannot be the equilibrium contract. ■

Proposition 1 *Assuming effort is not implemented in equilibrium, the only equilibrium contract* $\langle s_{1*}, s_{0*} \rangle$ *is characterized by the conditions*

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

and $q(x_1 - s_{1*}) + [1-q](x_0 - s_{0*}) = 0$. *The agent bears risk in equilibrium if* $\tilde{q} \neq q$.

Proof. Assuming that effort is not implemented in equilibrium, $\langle s_{1*}, s_{0*} \rangle$ characterized by $\frac{q}{1-q} = \frac{\tilde{q}}{1-\tilde{q}} \frac{u'(s_{1*})}{u'(s_{0*})}$ and zero expected profits for the principal maximizes the agent's perceived expected utility subject to non-negative profits for the offering principal. It can be shown that, because of the agent's risk aversion, this contract is the unique global maximum. Any other contract that does not implement effort and gives non-negative expected profits to the offering principal therefore yields strictly lower perceived expected utility to the agent. Conversely, any other contract that does not implement effort and yields the same perceived expected utility to the agent yields strictly negative expected profits for the offering principal. Thus, there is no profitable deviation from $\langle s_{1*}, s_{0*} \rangle$. ■

Proposition 2 *Assuming effort is implemented in equilibrium,* $\langle \bar{s}_1, \bar{s}_0 \rangle$ *is the only equilibrium contract if* $\frac{\tilde{q}+\tilde{v}}{1-(\tilde{q}+\tilde{v})} \frac{u'(\bar{s}_1)}{u'(\bar{s}_0)} \leq \frac{q+v}{1-(q+v)}$.

Proof. We need to show that there is no profitable deviation from this contract; i.e. that no other effort-implementing contract exists such that the agent receives higher expected utility, and the offering principal enjoys positive expected profits. This rules out contracts with higher payment to the agent in both success and failure, and those with lower payment to the agent in both events. The perceived expected utility for the agent under $\langle \bar{s}_1, \bar{s}_0 \rangle$ is

$$\tilde{\mathbb{E}}\left[u\left(\bar{s}_x\right) \mid e=1\right] = \left(\tilde{q}+\tilde{v}\right) u\left(\bar{s}_1\right) + \left[1-\left(\tilde{q}+\tilde{v}\right)\right] u\left(\bar{s}_0\right) - c.$$

A marginal change in payments that leaves the agent indifferent satisfies

$$\left(\tilde{q}+\tilde{v}\right) u'\left(\bar{s}_1\right) ds_1 + \left[1-\left(\tilde{q}+\tilde{v}\right)\right] u'\left(\bar{s}_0\right) ds_0 = 0$$

or

$$ds_0 = -\frac{\tilde{q}+\tilde{v}}{1-\left(\tilde{q}+\tilde{v}\right)} \frac{u'\left(\bar{s}_1\right)}{u'\left(\bar{s}_0\right)} ds_1$$

where necessarily $ds_1 > 0$ and $ds_0 < 0$, since the incentive compatibility constraint would be otherwise violated. Such a change implies that the agent would bear more risk under the new contract. Since the agent is risk-averse, he must be compensated with a higher expected payment, which is why the increase in s_1 must be more than actuarially fair *according to his beliefs.*

The change in expected profits for the offering principal from a marginal change in the payment structure that leaves the agent indifferent is

$$\begin{gathered} -\left(q+v\right) ds_1 - \left[1-\left(q+v\right)\right] ds_0 = \\ \left(-\left(q+v\right) + \left[1-\left(q+v\right)\right] \frac{\tilde{q}+\tilde{v}}{1-\left(\tilde{q}+\tilde{v}\right)} \frac{u'\left(\bar{s}_1\right)}{u'\left(\bar{s}_0\right)}\right) ds_1 \leq 0. \end{gathered}$$

If $\frac{\tilde{q}+\tilde{v}}{1-(\tilde{q}+\tilde{v})} \frac{u'(\bar{s}_1)}{u'(\bar{s}_0)} \leq \frac{q+v}{1-(q+v)}$, expected profits decrease for the principal offering the new contract that exposes the agent to more risk than $\langle \bar{s}_1, \bar{s}_0 \rangle$ does. It follows that there is no profitable marginal deviation from the $\langle \bar{s}_1, \bar{s}_0 \rangle$ contract. Given that $\frac{u'(s_1)}{u'(s_0)}$ falls as s_1 increases and s_0 decreases, discrete deviations from $\langle \bar{s}_1, \bar{s}_0 \rangle$ that implement effort and leave the agent indifferent also imply expected losses for the offering principal. Therefore, $\langle \bar{s}_1, \bar{s}_0 \rangle$ is the only equilibrium contract. ■

Proposition 3 *Assuming effort is implemented in equilibrium, if* $\frac{\tilde{q}+\tilde{v}}{1-(\tilde{q}+\tilde{v})} \frac{u'(\bar{s}_1)}{u'(\bar{s}_0)} > \frac{q+v}{1-(q+v)}$ *then* $\langle \bar{s}_1, \bar{s}_0 \rangle$ *is not an equilibrium contract. The only equilibrium contract* $\langle s_1^*, s_0^* \rangle$ *is characterized by*

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

and $\left(q+v\right)\left(x_1 - s_1^*\right) + \left[1-\left(q+v\right)\right]\left(x_0 - s_0^*\right) = 0.$ *The equilibrium contract has higher-powered incentives than necessary to implement effort.*

Proof. The proof is analogous to the proof of Proposition 1. Assuming that effort is implemented in equilibrium, $\langle s_1^*, s_0^* \rangle$ (uniquely) maximizes the agent's perceived utility subject to non-negative profits for the offering principal. It follows that any other contract that implements effort and yields the same perceived expected utility to the agent yields strictly negative expected profits for the offering principal. There is no profitable deviation from $\langle s_1^*, s_0^* \rangle$.

Given that by assumption $\frac{q+v}{1-(q+v)} < \frac{\tilde{q}+\tilde{v}}{1-(\tilde{q}+\tilde{v})} \frac{u'(\bar{s}_1)}{u'(\bar{s}_0)}$, since $u''(\cdot) < 0$, and expected profits for the principal are zero under both $\langle \bar{s}_1, \bar{s}_0 \rangle$ and $\langle s_1^*, s_0^* \rangle$, it follows that $s_1^* > \bar{s}_1$ and $s_0^* < \bar{s}_0$. By construction, $\tilde{v}\left(u\left(\bar{s}_1\right) - u\left(\bar{s}_0\right)\right) = c$, and thus

$$\tilde{v}\left(u\left(s_1^*\right) - u\left(s_0^*\right)\right) > c;$$

$\langle s_1^*, s_0^* \rangle$ has higher-powered incentives than necessary to implement effort. ■

Proposition 4 *Ceteris paribus, higher levels of overconfidence of either kind increase the likelihood that effort is implemented in equilibrium: if effort is implemented given agent 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. Overconfidence affects the implemented level of effort differently, depending on whether the incentive or the wager effect of overconfidence dominates.

First, consider changes in agent overconfidence regarding the value of effort.

If $\langle \bar{s}_1, \bar{s}_0 \rangle$ is the potential equilibrium contract that implements effort, the power of incentives decreases with overconfidence regarding the value of effort. Expected profits are zero in equilibrium, so a marginal increase in $\tilde{v}$ implies:

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

where $\frac{d\bar{s}_1}{d\tilde{v}} < 0$ and $\frac{d\bar{s}_0}{d\tilde{v}} > 0$. The effect of such a change in the agent's perceived utility when $\langle \bar{s}_1, \bar{s}_0 \rangle$ is the potential equilibrium contract that implements effort is

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

taking into account that $\langle \bar{s}_1, \bar{s}_0 \rangle$ is the potential equilibrium contract only if $\frac{\tilde{q}+\tilde{v}}{1-(\tilde{q}+\tilde{v})} \frac{u'(\bar{s}_1)}{u'(\bar{s}_0)} \leq \frac{q+v}{1-(q+v)}$.

If $\langle s_1^*, s_0^* \rangle$ is the potential equilibrium contract that implements effort, the envelope theorem implies that the change in the agent's perceived utility from a marginal increase in $\tilde{v}$ is

$$u\left(s_1^*\right) - u\left(s_0^*\right) > c > 0.$$

There is no change in the agent's perceived expected utility under $\langle s_{1*}, s_{0*} \rangle$ from a marginal increase in $\tilde{v}$. Therefore, if effort is implemented given agent beliefs $(\tilde{q}, \tilde{v})$, then effort will be implemented when his beliefs are $(\tilde{q}, \tilde{v}^*)$ for any $\tilde{v}^* \geq \tilde{v}$.

Consider now changes in agent overconfidence regarding the base probability of success.

If $\langle \bar{s}_1, \bar{s}_0 \rangle$ is the potential equilibrium contract that implements effort, the power of incentives is independent from overconfidence regarding the base probability of success. A marginal increase in $\tilde{q}$ implies a change in the agent's perceived expected utility of

$$u(\bar{s}_1) - u(\bar{s}_0) = c.$$

If $\langle s_1^*, s_0^* \rangle$ is the potential equilibrium contract that implements effort, the envelope theorem again implies that the change in the agent's perceived utility from a marginal increase in $\tilde{q}$ is

$$u(s_1^*) - u(s_0^*) > c.$$

The change in the agent's perceived expected utility following a marginal increase in $\tilde{q}$ under the potential equilibrium contract that does not implement effort is

$$u(s_{1*}) - u(s_{0*}) \leq c.$$

Therefore, if effort is implemented given agent beliefs $(\tilde{q}, \tilde{v})$, then effort will be implemented when his beliefs are $(\tilde{q}^*, \tilde{v})$ for any $\tilde{q}^* \geq \tilde{q}$. ■

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

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

where

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

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:

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

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:

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

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 regarding the value of effort

Recall that the solution to the agent's problem yields (3), which we can write as

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

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

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

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

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

or

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

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

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

Given that we are interested in the effect of overconfidence regarding the value of effort 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):

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

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

where

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

note $\frac{1}{\beta_1} - \frac{1}{\beta_0} > 0$.

The expression above shows that the power of incentives will be decreasing in overconfidence about the value of effort for some (sufficiently low) levels of overconfidence if the agent's action is very responsive to the power of incentives. This will be the case if the agent is very risk averse (as measured by $-\frac{u''(u(s_x))}{u'(u(s_x))}$, which slightly resembles the coefficient of absolute risk aversion), or if the increase in the marginal cost of effort is sufficiently low (as measured by $k = c''(e^*)$). The power of incentives of the optimal contract will be everywhere increasing in both kinds of overconfidence if the agent is sufficiently risk neutral, or if the disutility cost of effort is very convex.

References

Adrian, T. & Westerfield, M. (2007), Disagreement and learning in a dynamic contracting model. Working Paper FBE 04–07, Marshall Research Paper Series.

Akerlof, G. A. & Dickens, W. T. (1982), 'The economic consequences of cognitive dissonance', *American Economic Review* **72**(3), 307–319.

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

Bernardo, A. E. & 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. & Lovallo, D. (1999), 'Overconfidence and excess entry: An experimental approach', *American Economic Review* **89**(1), 306–318.

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

De Bondt, W. F. M. & Thaler, R. H. (1995), Financial decision-making in markets and firms: A behavioral perspective, *in* R. A. Jarrow, V. Maksimovic & W. T. Ziemba, eds, 'Handbooks in Operations Research and Management Science: Finance', Elsevier Science, pp. 385–410.

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

Eyster, E. & Rabin, M. (2005), 'Cursed equilibrium', *Econometrica* **73**(5), 1623–1672.

Fang, H. & Moscarini, G. (2005), 'Morale hazard', *Journal of Monetary Economics* **52**(4), 749–777.

Fiske, S. T. & Taylor, S. E. (1991), *Social Cognition*, second edn, McGraw-Hill, New York.

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

Gervais, S., Heaton, J. & Odean, T. (2003), Overconfidence, investment policy, and executive stock options. Center for Financial Research Working Paper No. 15–02.

Gervais, S. & Odean, T. (2001), 'Learning to be overconfident', *Review of Financial Studies* **14**(1), 1–27.

Goel, A. M. & Thakor, A. V. (2007), Overconfidence, leadership selection and corporate governance. Forthcoming in *Journal of Finance*.

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.

Kőszegi, B. (2005), Utility from anticipation and personal equilibrium. University of California at Berkeley, mimeo.

Kőszegi, B. & Rabin, M. (2006), 'A model of reference-dependent preferences', *Quarterly Journal of Economics* **121**(4), 1133–1165.

Koufopoulos, K. (2005), Asymmetric information, heterogeneity in risk perceptions and insurance: An explanation to a puzzle. Warwick Business School, mimeo.

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

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

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

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

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.

Prendergast, C. (2002), 'The tenuous trade-off between risk and incentives', *Journal of Political Economy* **110**(5), 1071–1102.

Santos-Pinto, L. (2007), Positive self-image and incentives in organizations. Forthcoming in *The Economic Journal*.

Smith, C. W. & Watts, R. L. (1992), 'The investment opportunity set and corporate financing, dividend, and compensation policies', *Journal of Financial Economics* **32**(3), 263–292.

Taylor, S. E. & 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.