Trust in Agency*

Ramon Casadesus-Masanell†

November 1999

Abstract

Existing models of the principal-agent relationship assume the agent works only under extrinsic incentives. However, many observed agency contracts take the form of a fixed payment. For such contracts to succeed, the principal must trust the agent to work in the absence of incentives. I show that agency fosters the advent of intrinsic motivation and trustworthy behavior. Three distinct motivational schemes are analyzed: norms, ethical standards, and altruism. I identify the conditions under which these mechanisms arise, and show how they promote trust. The analysis alters several important predictions of conventional models: total surplus is shared between principal and agent, the first best outcome ensues in highly uncertain environments, the principal is better off the more the agent is risk averse, and larger equilibrium extrinsic incentives need not be associated with larger effort or larger total surplus.

* I thank Daniel Spulber for his advice and encouragement throughout the project as well as Sandeep Baliga, Morton Kamien, Brian Uzzi, Nabil Al-Najjar, Scott Davis, Teresa Garcia Marco, Peter Klibanoff, Vernon Noronha, Emre Ozdenoren, Vicente Salas Fumas, and Joel Sandonis for helpful conversations. This is chapter 1 of my dissertation.
† Kellogg Graduate School of Management. Northwestern University. Departments of Management & Strategy and MEDS. [rcasades@nwu.edu]
1. Introduction

Trust is a critical element that facilitates social and economic interaction. Agency contracts, in particular, depend fundamentally on trust between principal and agent. In Law, agency is defined as a fiduciary relationship, that is, one that is founded on trust. However, economic models of agency have tended to assume an absence of trust, focusing instead on moral hazard and adverse selection. To evaluate the role of trust in agency contracts, I examine a moral hazard model of the principal-agent relationship. The model considers three significant bases for trust: norms, ethical standards, and altruism.

I examine the circumstances in which it is in the agent’s best interest to develop intrinsic motivation or, in Frank’s [9] terms, to “acquire a conscience.” By intrinsic motivation I mean endogenously-determined enticement to work based on preferences, without the intervention of external punishments and rewards. I show that such motivation translates into trust: the offering of low-powered incentive schemes and development of a willingness to contract in settings where extrinsic incentives cannot be given. Intrinsic motivation has been largely overlooked by economists because of the focus on behavioral variations taking preferences as given. Important work by Kandel and Lazear [13] and Rotemberg [21] studies intrinsic motivation in the context of partnership games. I explore some of the implications of their work for the design of efficient agency contracts.

The standard model of the principal-agent relationship shows that when risk aversion is high and when the available information signal on effort has high variance, extrinsic incentives result in substantial risk bearing, hampering the principal’s ability to motivate work. In addition, it may be costly for the principal to design and implement incentive schemes. If incentives need be absent in a setting where effort is unobservable, why should the principal expect any work to be done? Homo economicus will shirk and trade will break down. An effort-averse agent may choose to work under the fear of future punishments for low performance. This approach has been extensively studied in the context of repeated games and I do not consider it in what follows so as to focus on intrinsic preferences. Understanding the advent of intrinsic motivation sheds some light on David Kreps’s [15] observation that larger monetary incentives appear to be empirically associated with lower intrinsic motivation, lower effort, and lower total surplus. The standard model is inconsistent with Kreps’s
observation because larger equilibrium extrinsic incentives are associated with larger total surplus.

I begin by considering the role of norms in agency contracts. Social psychologists have defined a norm as a rule or guide to behavior determining what is appropriate or inappropriate. When the agent chooses an action different from the norm, disagreeable feelings develop if the observable realization of the random information signal amplifies the agent’s deviation from the norm. I interpret such distress as a feeling of shame. Norms affect the agent ex post because the extent of the psychological distress depends on what the realized signal is, and this is known only after uncertainty has unfolded. The most the agent can do ex ante by his choice of action is to minimize the expected psychological cost. Because the agent worries about the realized information signal misrepresenting his choice of action relative to the norm, the pressure he feels to observe norms increases with the variability of the signal and his degree of risk aversion. When the level of perceived risk is low, norms are “high” (larger than the first best action) and the principal’s participation constraint binds. Moreover, since the pressure to abide by the norm is low when risk is low, the agent feels little motivation to abide by the norm and the principal may be better off by not trusting him. As perceived risk increases, the equilibrium norm gets closer to the first best action and principal and agent share total surplus. Also, since pressure to observe the norm increases, the principal is confident that the agent will honor contracts that ask for large effort while providing low-powered incentives and a large fixed payment; thus, she trusts him. I show that the equilibrium norm approaches the first best action as the informativeness of the signal on effort deteriorates. Therefore, where traditional analyses predict a breakdown of the relationship, endogenous norms deliver the first best outcome. I conclude that appealing to responsibility and committing to less precise observation technologies may result in better outcomes.

Next, I examine the role of ethical standards as a source of trust in agency. By an ethical standard I mean a level of effort that is considered “appropriate behavior.” The agent feels distress when his unobservable choice of action diverges from the ethical standard. This is interpreted as a feeling of guilt. Ethical standards work ex ante because the extent of the anguish depends on the action taken, and not on the value the information signal takes ex
post. High ethical standards result in intrinsic motivation because the closer is effort to the standard the less the psychological cost borne by the agent. Therefore, to implement any given action, the principal needs to offer lower extrinsic incentives than if the agent did not have ethical standards. Also, in order to persuade an agent with ethical standards to enter the agency relationship, the principal needs to offer a larger fixed payment because on top of the physical cost of any given action, there is the psychological cost imposed by the standard. I show that it is in the agent’s best interest to develop pressure to abide by the standard and that the equilibrium ethical standard corresponds to the first best action. Principal and agent share total surplus even if the principal has all bargaining power at the contracting stage. Contrary to the standard model, when the level of risk perceived by the agent goes to infinity, the equilibrium action converges to half of the first best action and total surplus does it to 3/4 of the first best total surplus, even as extrinsic incentives vanish. An agent who has developed pressure to abide by the standard is trustworthy and is always trusted. Trust and trustworthiness result in larger surplus as compared to the no-trust situation. Finally, if the principal commits to a contract before the ethical standard develops, the no-trust outcome takes place.

Finally, I study altruism in agency. An agent is altruistic when the principal’s well-being enters positively in his utility function. Altruism has a double effect on the agent’s perception as compared to the case in which he is non-altruistic. First, he senses each action as being less costly because he derives satisfaction from seeing the principal benefiting from the action. Second, when the principal’s gross benefit function has randomness, the agent also bears additional risk: he is worried that the principal will feel unhappy if the random outcome of his action turns out to be low. These two effects result in the agent being intrinsically motivated to work. As a consequence, to implement any given effort level, the principal can now offer lower powered incentives. Also, to persuade the agent to enter the relationship, she needs to offer a larger fixed wage to compensate for the additional risk bearing. Thus, altruism results in the principal trusting the agent. Is it ever in the agent’s best interest to develop altruism? Low-powered extrinsic incentives may be desirable because they result in lower risk bearing; nevertheless, they also result in lower total payment. In low risk environments (those in which the signal informing of the agent’s choice of action has
low variance or the agent is only slightly risk averse) risk bearing is already low, therefore the agent has little incentive to become altruistic. As the level of perceived risk grows larger, the desirability of lower powered incentives and, thus, of altruism, increases. I show that there is a threshold level of perceived risk above which the agent chooses to be altruistic. However, in environments where extrinsic incentives are too costly to provide and altruism would be the most socially desirable, the agent will not develop altruism.

One important conclusion from all three models is that exposing the agent to risk may be desirable to both principal and agent. Norms are most advantageous to the principal when perceived risk is high; ethical standards also result in better outcomes under large risk; altruism is maximum in medium risk environments, low risk promotes the emergence of spite. This suggests that organizational practices that expose workers to risk will result in better outcomes than if the job is narrowly defined and under conditions of certainty. Participative management, by increasing the workers’ awareness to the risks of operation, will result in the development of welfare enhancing norms, ethical standards, and altruism.

Even though Kenneth Arrow [3] observed over a quarter of a century ago that “[t]rust is an important lubricant of a social system,” only very recently have economists become actively involved in research on trust. Fisman and Khanna [8], Knack and Keefer [14], La Porta, Lopes-de-Silanes, Shleifer, and Vishny [16], Portales, Ricart-i-Costa, and Rosanas [17], Salas [22], Spagnolo [25], and Williamson [26] study trust but do not consider its function in a principal-agent setting. The role of trust in agency had been previously noted by Arrow [2] and Rosen [20]. Arrow considers trust between principal and agent as a necessary condition for economic prosperity: “One of the characteristics of a successful economic system is that the relations of trust and confidence between principal and agent are sufficiently strong so that the agent will not cheat even though it may be ‘rational economic behavior’ to do so. The lack of such confidence has certainly been adduced by many writers as one cause of economic backwardness,” [emphasis added]. Rosen notes that “[individual owners] must place a certain amount of trust in a management team to take proper actions on their behalf. Herein lies the agency problem,” [emphasis added]. Neither Arrow nor Rosen presents a formal model to make his claims precise.
The remainder of the paper is organized as follows. Section 2 presents the basic agency model and defines trust and trustworthiness. Sections 3, 4 and 5 examine norms, ethical standards, and altruism. Section 6 compares the three motivational schemes. Section 7 presents conclusions.

2. The Model

I distinguish between material payoffs (the true utility function, or the object self) and behavior payoffs (the utility function used in acting, or the acting self). Viewing the self as an entity of parts with some components making use of others is common in sociology, psychology, and social psychology.1 In The Theory of Moral Sentiments in 1753, Adam Smith distinguished between the object self, the acting self, and the “judge,” who evaluates the agent’s and others’ actions. This notion also has been used in economics by Akerlof [1], Frank [9], Raub [19], Rabin [18], and Rotemberg [21], among others.

Material payoffs refers to the standard agent as in most of the principal-agent literature, that is, an unsocialized, effort-averse, and selfish individual. The agent’s material payoffs are taken as given. In contrast, behavior payoffs may display norms, ethical standards, or altruism. Rational norms, ethical standards, and altruism are endogenously generated within the model.

Principal and agent play the following two-stage game introduced first by Rotemberg [21]. In stage 1, the agent molds his behavior payoffs to serve best his object self. That is, the agent uses his selfish utility function to evaluate the desirability of norms, ethical standards, and altruism.2 In the case of norms he decides the norm; with ethical standards, he chooses whether to develop pressure to abide by the standard; in the case of altruism, he chooses how much to empathize with the principal. Just as in Rotemberg [21], I assume that once the agent has chosen his behavior payoffs, he is bound to act according to them. In other words, the agent credibly commits to observable norms, ethical standards, and altruism. Credibility requires that it be difficult for an agent who is bound by norms, ethical standards or altruism

---

1See Coleman [5, chapter 19] for extensive references.
2For a brief survey of the literature that features this opportunistic change in tastes, see Rotemberg [21, footnote 1].
to imitate the signals that those truly experiencing these emotions display.³

In stage 2, a fully standard agency relationship unfolds. The principal offers a contract to the agent; the agent, given his behavior payoffs, decides whether to enter the relationship and, if he does, then he chooses how much effort to exert. The principal cannot observe the agent’s effort choice. After an observable and verifiable signal on effort has been realized, a monetary transfer takes place.

Rotemberg [21] distinguishes a number of interpretations for the first stage of this game. First, a selfish inner self relinquishes control of actions to the outer self. The inner self can mold the preferences that guide the outer self choices. The inner self can make the outer self altruistic, bound by ethical standards or norms. Such feelings become genuine because the inner self can only change the outer self’s preferences slowly. Second, if emotional reactions are guided by genes, natural selection might favor the reproduction of individuals whose emotions change in a self-interested way. Finally, natural selection favors those genes that lead people to imitate the behavior of individuals that appear successful. If people appear successful when their material payoffs are high and the parameter values of successful individuals can be inferred from their behavior, then people will be led to choose parameters in a way that maximizes expected material payoffs.

2.1. Material payoffs or the “object self”

This section presents the standard agency preferences. Assume the agent is selfish and effort-averse. Let \( a \) be the action taken by the agent and \( a + x \) the observable informative signal on effort. Assume for simplicity that the shock \( x \) is normally distributed with zero mean and variance \( \sigma^2 \). Let \( \alpha \) be a commission rate and \( \beta \) a fixed payment. The pair \( \langle \alpha, \beta \rangle \) is referred to as an incentive schedule. A contract \( \langle \alpha, \beta, a \rangle \) is an effort level, \( a \), together with an incentive schedule. The principal offers linear wage \( \alpha (a + x) + \beta \) and expects the agent

³For example, Frank [9, 594-5] points out that “[a] strategically important emotion can be communicated credibly only if it is accompanied by a signal that is at least partially insulated from direct control. Many observable physical symptoms of emotional arousal satisfy this requirement. Posture, the rate of respiration, the pitch and timber of the voice, perspiration, facial muscle tone and expression, movement of the eyes, and a host of other readily observable physical symptoms vary systematically with a person’s affective condition.”
to choose action $a$. The agent is paid $\alpha (a + x) + \beta$ after signal $a + x$ has been realized.\footnote{As shown in Holmstrom and Milgrom [11], the assumptions of a linear incentive schedule together with the agent choosing effort once and for all without regard to the arrival of performance information over time, is without loss of generality if one interprets the model as the agent choosing effort continuously over the time interval $[0, 1]$ to control the drift parameter of a stationary brownian motion and the agent can observe his accumulated performance before acting.}

The agent’s cost of undertaking action $a$ measured in dollar terms\footnote{The results remain virtually unchanged if one assumes a more general cost function $C$ with the usual properties: (1) $C' \geq 0$; (2) $C'' > 0$. To perform comparative statics, it is convenient to assume further that $C''' \geq 0$.} is $C(a) = a^2$. When the agent faces incentive schedule $(\alpha, \beta)$ and takes action $a$, he derives a monetary rent of $\alpha (a + x) + \beta - a^2$. Assuming that the agent has a von Neumann-Morgenstern utility function with constant absolute risk aversion, the corresponding utility is

$$u = - \exp \left\{ -r \left( \alpha (a + x) + \beta - a^2 \right) \right\}, \quad (2.1)$$

where $r > 0$ is the coefficient of absolute risk aversion.

The greater the variability of the shock, $\sigma^2$, the greater the risk the agent perceives for any given level of risk aversion, $r$. Also, the larger is risk aversion, $r$, the more sensitive the agent is to the variability of the shock $x$. Therefore, $r\sigma^2$ represents the level of perceived risk. To simplify notation, let $k \equiv r\sigma^2$. The agent’s certainty equivalent of action $a$ is

$$U = \alpha a + \beta - a^2 - \frac{1}{2}k\alpha^2. \quad (2.2)$$

This is the expected payment given $a$, $\alpha a + \beta$, minus the “physical” cost of the action, $a^2$, minus the cost of risk bearing, $\frac{1}{2}k\alpha^2$. Note that absent extrinsic incentives ($\alpha = 0$), the agent is better off not working. Thus, material preferences correspond to an individual devoid of any form of intrinsic motivation, someone who cannot be expected to work in the absence of extrinsic incentives – homo economicus in its purest form.

Let $\tilde{B}(a, x) = a + x$ be the gross benefit to the principal when the information signal takes value $a + x$.\footnote{The results remain the same if one assumes a more general gross benefit function $\tilde{B}(a, x) : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$ with the following properties: (1) $x$ enters linearly so that $E\tilde{B}(a, x) = B(a)$; (2) $B' > 0$; (3) $B'' \leq 0$.} The principal is assumed to be risk neutral. Thus, her certainty equivalent
of action $a$ is,

$$V = a - \alpha a - \beta.$$  \hfill (2.3)

### 2.2. A Benchmark: The Standard Agency Contract

An incentive-efficient linear contract maximizes joint surplus $U + V$ subject to the incentive compatibility constraint. Thus, a contract is incentive-efficient if it is a solution to

$$\max_{\alpha, \beta} U + V \text{ subject to } a \in \arg \max_a U.$$ \hfill (2.4)

Normalizing the agent’s reservation certainty equivalent to zero, $\beta$ is set to allocate total surplus between principal and agent. Since the principal makes a take-it-or-leave-it offer, she sets the fixed component of wage, $\beta$, so that the agent is just willing to participate. Thus,

$$\beta = -a\alpha + a^2 + \frac{1}{2}ka^2.$$

The equilibrium contract is

$$\alpha^S = \frac{1}{1 + 2k}, \quad \beta^S = \frac{2k - 1}{4(1 + 2k)^2}, \quad \text{and} \quad a^S = \frac{1}{2(1 + 2k)}.$$

The resulting certainty equivalents are

$$U(\alpha^S, \beta^S, a^S) = 0 \quad \text{and} \quad V(\alpha^S, \beta^S, a^S) = \frac{1}{4(1 + 2k)}.$$

We can plot the equilibrium values as the level of perceived risk, $k$, varies. This will become useful later when comparing the models; see Figure 1.
Figure 1: Equilibrium in the standard model as a function of perceived risk $k$.

The first best outcome $U^* + V^* = \frac{1}{4}$ is obtained by maximizing total surplus under perfect observability. The first best action is $a^* = \frac{1}{2}$. The equilibrium outcome under imperfect information is inefficient. To provide incentives for the agent to work, the principal must make him bear risk which reduces total welfare. The degree of inefficiency is

$$ (U^* + V^*) - (U (\alpha^S, \beta^S, a^S) + V (\alpha^S, \beta^S, a^S)) = \frac{k}{2 + 4k}, $$

(2.5)

an increasing function of perceived risk, $k$.

The standard model of agency makes a number of stylized predictions. First, the agent is left with no rents. Second, when perceived risk $k$ grows to infinity, the agent cannot be induced to work and thus the relationship necessarily breaks down. It is important to note that if the agent’s cost function had $C'(0) > 0$, then a sufficiently large $k$ but strictly less than $\infty$ would result in the equilibrium action being zero.\footnote{The first order condition in the agent’s maximization problem is $\alpha^S = C'(a^S)$; thus, for $a^S$ to be set larger than zero it is necessary that $\alpha^S > C'(0) > 0$. Since $\alpha^S$ approaches zero as $k$ tends to infinity, a sufficiently large $k < \infty$ is enough to have $a^S = 0$ and thus for the relationship to break down. For example, if $C(a) = (a + m)^2 - m^2$ with $m \in [0, \frac{1}{2}]$, the agent will only undertake positive effort if $\alpha$ is larger than $2m$. But the equilibrium $\alpha$ (for an interior solution) is $\alpha^S = \frac{1}{1+2k}$. Thus as long as $k \geq \frac{1-2m}{4m}$, the relationship...} Hence, the implication that the
agency relationship breaks down when the level of perceived risk is large is more general than what it may seem given the assumed cost function. Third, the more risk averse is the agent, the less utility the principal realizes. Finally, since $a^S = \frac{\alpha^S}{2}$ and $U(\alpha^S, \beta^S, a^S) + V(\alpha^S, \beta^S, a^S) = \frac{\alpha^S}{4}$, larger equilibrium extrinsic incentives (resulting from lower $k$) must necessarily be associated with a larger action and larger total surplus.

**2.3. Trust**

The inefficiency captured by the standard model of agency (eq. 2.5) is due to moral hazard: insufficient extrinsic incentives coupled with unobservability (or, more generally, unverifiability) result in the agent under-performing. Moral hazard forces the principal to provide extrinsic incentives that impose welfare-reducing risk bearing. Intuitively, larger total surplus would be obtained if the agent was ready to honor contracts that asked for large effort while offering low-powered incentives. The agent would then be compensated through the fixed component of salary. This would avoid risk bearing and result in larger gains.

Because the agent shirks when guided by his material payoffs unless appropriate extrinsic incentives are granted, I refer to his material payoffs as untrustworthy preferences. Furthermore, since $\langle \alpha^S, \beta^S, a^S \rangle$ is the optimal contract when the agent acts to satisfy his untrustworthy preferences, I refer to it as the no-trust contract. Therefore, I use the standard economic agency model as the no-trust benchmark. What is trust in the context of the principal-agent relationship?

**Definition 1.** The agent is **trustworthy** if he is willing to work in the absence of extrinsic incentives. The principal trusts the agent if she offers him a contract $\langle \alpha, \beta, a \rangle$ with $\alpha \leq \alpha^S$ and $\beta \geq \beta^S$ with at least one of the inequalities strict, and $a \geq a^S$. All contracts satisfying these conditions are referred to as trust contracts.

The principal trusts the agent if she offers him lower powered incentives while asking for at least as much effort as under the no-trust contract. Notice that trust is defined as behavior and not as the cause of behavior.8 In other words, I call trust the offering of a

---

8For a critique of Williamson’s [26] notion of calculativeness based on the grounds that he does not distinguish between trust as behavior and trust as cause of behavior, see Craswell [6].
certain contract. I do not say that such a contract is offered because the principal trusts the agent. I will examine reasons that such behavior takes place. Note also that nothing in the definition implies that when the principal trusts the agent and trust is honored, the resulting total surplus is larger than when the agent has untrustworthy preferences. In fact, I will show that there are circumstances in which trust does not necessarily translate into larger total surplus.

It is useful to compare my definition of trust to that offered in the sociology literature. According to sociologist Mark Granovetter [10], “the most common definition of ‘trust’ is precisely the confidence that another will not take advantage of you despite clear incentives to do so, even in ‘end-game’.” In the present context, the “confidence that another will not take advantage of you” is captured in that when the principal trusts the agent, if the agent acts to maximize his material payoffs, then the principal is worse off than if she had offered the no-trust contract. Furthermore, and as I will show below, the principal may have been better off by staying out of the relationship. “[D]espite clear incentives to do so” is captured in that the object self has always an incentive to breach a trust contract. Finally, “even in ‘end-game’,” is reflected in that the model is static and compliance cannot rely on self-enforcement of the type considered in the repeated games literature.

Trust contracts are not incentive compatible. However, as Partha Dasgupta [7] points out, “you do not trust a person (or an agency) to do something merely because he says he will do it. You trust him only because, knowing what you know of his disposition, his available options and their consequences, his ability and so forth, you expect that he will choose to do it.” This apparent contradiction is resolved by the distinction between the object and the acting self. Trust contracts are not incentive compatible to the agent’s object self, but they need be so to the agent’s acting self. This naturally restricts the set of admissible trust contracts. This distinction also helps explain Arrow’s [2] observation that “[o]ne of the characteristics of a successful economic system is that the relations of trust and confidence between principal and agent are sufficiently strong so that the agent will not cheat even though it may be ‘rational economic behavior’ to do so.” When facing a trust contract, it is at the same time rational for the object self and irrational for the acting self to cheat. I turn now to the study of equilibrium acting self preferences and trust under norms, ethical
standards, and altruism.

3. Norms

The first internal motivational scheme that I consider as a basis for trust in agency are norms. Social psychologists have defined a norm as a “shared rule or guide to behaviour that is appropriate or inappropriate.” The norm disciplines behavior through social (or external) pressure: the agent experiences disagreeable feelings when, having chosen an action different than the norm, outsiders (such as the principal or third parties) observe realizations of the random information signal that amplify the agent’s deviation.

To model norms, I let the agent’s acting self preferences be given by

\[ u = -\exp \left\{ -r (\alpha (a + x) + \beta - a^2 + \gamma (\nu - a) x) \right\}. \]

These are material preferences together with

\[ \gamma (\nu - a) x. \]  \hspace{1cm} (3.1)

\( \nu \) is the norm and \( \gamma \geq 0 \) a parameter correlated with the pressure felt by the agent to comply with the norm.

Expression (3.1) says that the agent can avoid potential embarrassment by choosing the action to be the norm, that is, by doing the “right” thing. If he chooses an action different from the norm, then if the random shock \( x \) amplifies his misbehavior, then he feels distressed. I interpret this as a feeling of shame. Thus, if \( a < \nu \) and \( x < 0 \), then \( a + x < a < \nu \) and his misbehavior is evidenced by the random shock \( x \). Also, had he chosen an action closer to the norm, the distance between the realized signal and the norm would have been smaller. On the other hand, if \( a < \nu \) and \( x > 0 \), then \( a < a + x \) and his misbehavior is hidden. This hiding makes the agent feel relieved and thus somewhat happy. If the shock is large enough so that \( a + x > \nu \), then not only there is hiding of the agent’s wrongdoing but also, had he chosen an action closer to the norm, the realized signal would have been even further away

\(^9\)Brown [4, 49].

13
from it, a further reason to feel relief. A similar reasoning applies to the case in which the action is larger than the norm.

Because the agent takes his action before uncertainty unfolds, it is instructive to look at (3.1) from an ex-ante (before the action is taken) point of view. The agent’s certainty equivalent of (3.1) is

$$-\frac{1}{2} k\gamma^2 (\nu - a)^2. \tag{3.2}$$

Unless the action is chosen to be the norm, norms impose a cost on the agent, even if (3.1) may be positive ex-post. Thus, abiding by the norm \(a = \nu\) minimizes expected psychological cost. Equation (3.2) very much resembles Kandel and Lazear’s [13] modeling of norms in partnership games.

The term

$$\frac{1}{2} k\gamma^2 \tag{3.3}$$

is the pressure felt by the agent to observe the norm. Of course, when (3.3) is zero, norms have no effect on the agent’s behavior: the agent’s acting and material payoffs coincide. Because \(k \equiv r\sigma^2\), pressure depends on the level of risk aversion as well as on how noisy the information signal is. When the agent perceives high risk, he feels very pressured to choose his action close to the norm.

The agent’s certainty equivalent of action \(a\) is

$$U^N = \alpha a + \beta - a^2 - \frac{1}{2} k (\alpha + \gamma (\nu - a))^2. \tag{3.4}$$

Notice that the norm enters risk bearing only. A large norm induces work because by choosing

---

10Note the asymmetry in the way a positive and a negative shock enters the agent’s utility: when \(x = 0\) the utility evaluation of (3.1) gives \(-\exp (-r (0)) = -1\). The maximum relief the agent can feel occurs when \(x = +\infty\) and this results in an increase of utility of 1 util. On the other hand, maximum distress occurs when \(x = -\infty\) and this results in a decrease in utility of \(-\infty\) utils. Therefore, the agent is much more concerned about getting a shock that exacerbates his wrongdoing than one that hides it.

11That is,

$$\int_{-\infty}^{\infty} -\exp (-r (\gamma (\nu - a) x)) \frac{\exp \left(-\frac{1}{2} \frac{x^2}{\sigma^2}\right)}{\sqrt{2\pi}} dx = -\exp \left(-r \left(-\frac{1}{2} k\gamma^2 (\nu - a)^2\right)\right).$$
an action close to the norm, the agent is able to avoid risk. I now study equilibrium norms taking \( \gamma \) as given. In subsection 3.4 I endogeneize \( \gamma \) and show that it is always in the agent’s best interest to set it strictly positive.

### 3.1. Rational norms

The optimal action as a function of extrinsic incentives and the norm is

\[
a (\alpha (\nu), \nu) = \frac{\alpha + k\gamma \alpha + k\gamma^2 \nu}{2 + k\gamma^2}.
\]  

(3.5)

This is increasing in both the norm and extrinsic incentives. Thus, norms result in intrinsic motivation. Intuitively, the larger is the norm, the larger is the action that leaves the agent at any given level of risk bearing.

The optimal commission rate as a function of the norm is

\[
\alpha (\nu) = \frac{k\gamma - 2k\gamma \nu + 1}{2k + 1}.
\]  

(3.6)

The fixed component of wage, \( \beta (\nu) \), is chosen so that the agent is willing to participate given his acting self preferences. Notice that the larger is the norm, the lower is the principal’s willingness to provide extrinsic incentives. Therefore, there is a trade off between intrinsic motivation and extrinsic incentives. To understand this, note that the first best action \( a^* \) is a constant for all values of the norm. In fact, \( a^* - a (\alpha (\nu), \nu) \) decreases with \( \nu \). Consequently, the larger is the norm, the less are the extrinsic incentives required to have the agent choose an action close to first best.

The agent selects the norm so as to maximize his material payoffs subject to the principal’s participation constraint. The equilibrium norm is the solution to

\[
\max_{\nu} U \quad \text{subject to} \quad V \geq 0.
\]  

(3.7)

Because in the model of the principal-agent relationship outsiders are left unmodeled, the equilibrium norm is a personal (as opposed to a social) norm.
Proposition 1. For each \( \gamma \) there is a number \( k(\gamma) \) such that if \( k \leq k(\gamma) \), the equilibrium norm is

\[
\nu = \frac{1}{2\gamma} \left( \gamma - 2 + \left( \frac{(2k + 1)(2 + k^2)}{k} \right)^{\frac{1}{2}} \right),
\]

and the principal’s participation constraint is binding. When \( k > k(\gamma) \), the equilibrium norm is

\[
\nu = \frac{1}{2} \left( \frac{k(2 + k^2)^2(1 + k\gamma) + (2 - \gamma)(1 + k\gamma)^2}{\gamma \left( k^2(2 + k^2)^2 - (1 + k\gamma)^2 \right)^{\frac{1}{2}}} \right),
\]

and the principal’s participation constraint is not binding.

The equilibrium contract is

\[
\alpha^N = \begin{cases} 
\left( 1 - \left( \frac{k(2 + k^2)}{2k + 1} \right)^{\frac{1}{2}} \right) & \text{if } k \leq k(\gamma) \\
\left( \frac{(1 + k\gamma)^2}{k^2(2 + k^2)^2 - (1 + k\gamma)^2} \right)^{\frac{1}{2}} & \text{if } k > k(\gamma)
\end{cases}
\]

\[
\beta^N = \begin{cases} 
\left( \frac{1}{2} \left( \frac{k\gamma - 2k}{2k + 1} + \left( \frac{k(2 + k^2)}{2k + 1} \right)^{\frac{1}{2}} \right) \right) & \text{if } k \leq k(\gamma) \\
\left( \frac{1}{4} \left( \frac{(1 + k\gamma)^2}{k^2(2 + k^2)^2 - (1 + k\gamma)^2} \right)^{\frac{1}{2}} \right) & \text{if } k > k(\gamma)
\end{cases}
\]

\[
a^N = \begin{cases} 
\left( \frac{1}{2} \left( 1 + (\gamma - 2) \left( \frac{k}{(2k + 1)(2 + k^2)} \right)^{\frac{1}{2}} \right) \right) & \text{if } k \leq k(\gamma) \\
\left( \frac{1}{2} \frac{k^2\gamma(2 + k^2)^2(1 + k\gamma) - (1 + k\gamma)^2}{k^2(2 + k^2)^2 - (1 + k\gamma)^2} \right)^{\frac{1}{2}} & \text{if } k > k(\gamma)
\end{cases}
\]

Assuming \( \gamma = 1 \), the cutoff value is \( k(\gamma) \approx 1.76 \). We can plot the equilibrium values as functions of the level of perceived risk \( k \equiv \rho \sigma^2 \); see Figure 2. The solid line corresponds to the equilibrium in the norms model. The dotted line portrays the equilibrium with material preferences.
Figure 2: Effects of endogenous norms on the equilibrium contract and outcome as the level of perceived risk, $k$, varies.

The equilibrium norm and action have the following properties,

**Proposition 2.** $\nu > a^*$ and $a^N < \nu$.

Thus, the norm is larger than the first best action and the equilibrium action is less than the norm. The norm is chosen by the agent, hence a “relaxed” norm that does not induce much work is always available. Why does the agent end up choosing a large norm that constrains him to work hard? If the norm is low, then the principal provides incentives that make the agent choose an action larger than the norm. This results in $\gamma (\nu - a) < 0$ and thus in low risk bearing to the acting self (because $\alpha$ and $\gamma (\nu - a)$ cancel out to some extent). $\beta$ is then set low and the object self is worse off than if norms were absent. Therefore, the agent is always better off by setting the norm above the first best action.

It is important to note that even if the norm is always “demanding” in the sense that $\nu > a^*$, when $k > k (\gamma)$ it is not as demanding so as to leave the principal with no rents.
When the level of perceived risk \( k \) is large enough \( (k > k(\gamma)) \), both principal and agent share total surplus and as \( k \to \infty \) it is the agent’s rents that dissipate, whereas the principal’s increase. Intuitively, because the principal can always choose \( \alpha < 0 \), there is a maximum level of risk bearing to which the agent can effectively commit by his choice of the norm. This level vanishes as perceived risk increases.

The second part of proposition 2 says that the equilibrium action is always less than the equilibrium norm. Since the equilibrium norm is larger than the first best action, the principal sets extrinsic incentives so that the agent chooses an action below the norm.

Sen [23] observes that “[n]o society would be viable without some norms and rules of conduct. Such norms and rules are necessary for viability exactly in fields where strictly economic incentives are absent and cannot be created.” In the standard model of agency, when the ability to provide incentives deteriorates, the equilibrium action and total gains from the relationship tend to zero. The following proposition shows that endogenous norms arise when Sen regards them as most necessary. Moreover, the equilibrium norm and the resulting total surplus tend to the first best.

**Proposition 3.** \( \lim_{k \to \infty} \alpha^N = \lim_{k \to \infty} \nu = a^* \), \( \lim_{k \to \infty} \alpha^N = 0 \), \( \lim_{k \to \infty} V(\alpha^N, \beta^N, a^N) = V^* \), and \( \lim_{k \to \infty} U(\alpha^N, \beta^N, a^N) = U^* \).

Note that these limits are independent of \( \gamma \). Two points demand explanation. First, why does the equilibrium action converge to the norm as perceived risk grows? As perceived risk grows, the agent’s worry of being misrepresented by the information signal increases. In the limit, this concern is so present that he is better off by not taking any chances.

Second, why does the equilibrium norm converge to the first best action? Suppose that the level of perceived risk \( k \) is very large but that the agent chooses the norm to be larger than the first best action. Then, it is best for the principal to reduce incentives by setting \( \alpha \) close to but smaller than 0. This makes the acting self choose \( a \) close to but strictly less than the norm \( \nu \) to compensate for the risk bearing caused by \( \alpha \). While this choice is optimal for the acting self, it has shattering consequences for the object self, because now he bears huge risk (because \( k \) is large). Hence, as perceived risk grows large, the agent chooses the norm close to the first best action.
Contrary to the existing literature, I conclude that large risk is preferable to the principal: it may be best for the principal to refuse to monitor the agent, use less perfect observational technologies, appeal to responsibility, and provide flat incentive schemes.

3.2. Trust

Absent extrinsic incentives, an agent who has developed norm \( \nu \), is willing to work up to

\[
a = \frac{k \gamma^2 \nu}{2 + k \gamma^2}.
\]

Hence, equilibrium norms result in the agent being trustworthy. Of course, more effort can be induced if extrinsic incentives are provided. Is the principal always willing to trust the agent?

**Proposition 4.** *The agent is trusted if he perceives enough risk.*

Because pressure to abide by the norm motivates the agent to work, the principal provides lower powered extrinsic incentives \( (\alpha^N < \alpha^S) \). Nevertheless, when the agent perceives low risk, if \( \gamma \) is also low, then he feels low pressure to observe the norm and the mere presence of the norm is not sufficient for the principal to be willing to trust him. Also when perceived risk is low, total surplus under norms is less than under standard preferences.\(^{12}\)

Finally,

**Proposition 5.** *If the agent was to act to maximize his material payoffs, the principal would have been better off by not entering the relationship.*

\(^{12}\)In fact, if \( \gamma < 2 \) and perceived risk \( k \) is low, then \( a^N < a^S \), and the equilibrium contract under norms does not satisfy one of the requirements for it to be a trust contract. To see this, note that at \( k = 0 \), \( a^N = a^S = \frac{1}{2} \). Also, \( \lim_{k \to 0^+} \frac{da^N}{dk} = -\text{sign} \left( -\frac{1}{2} \gamma + 1 \right) \) and \( \lim_{k \to 0^+} \frac{da^S}{dk} = -1 \). Thus, when \( \gamma < 2 \), \( a^S > a^N \) for \( k \) close enough to zero.

Also, note that \( (U + V)^N|_{k=0} = (U + V)^S|_{k=0} = \frac{1}{4} \) and

\[
\lim_{k \to 0^+} \frac{d(U + V)^N}{dk} = -\frac{1}{2} \left( 1 + \frac{(\gamma-2)^2}{4} \right) \leq -\frac{1}{2} = \lim_{k \to 0^+} \frac{d(U + V)^S}{dk}.
\]

Proposition 3 guarantees that with enough risk, total surplus becomes larger under norms.
The principal trusts the agent not only in the sense that she offers lower powered incentives and larger fixed payment while expecting large effort, but also in the sense that if the agent was to act according to his material payoffs, she would have been better off staying out of the relationship.

3.3. Committing to a contract

Would the principal ever wish to commit to a contract before the agent chooses the norm? Commitment amounts to changing the order of the moves in the game. The principal chooses a contract and then the agent chooses the norm and action. It is easy to see that with commitment, the no-trust outcome takes place. Thus, the principal is left with surplus

\[ V = \frac{1}{4(2k + 1)}. \] (3.8)

Since when risk is low, endogenous norms result in the principal realizing no surplus, she is better off by committing to a contract. As \( k \) grows, equation (3.8) is reduced, whereas \( V(\alpha^N, \beta^N, a^N) \) approaches the first best value \( V^* \), therefore, commitment is inferior to non-commitment when the agent’s perception of risk is high.$^{13}$

3.4. Rational Social Pressure

If we add a stage zero in which the agent chooses parameter \( \gamma \), then he always chooses it to be strictly positive.

**Proposition 6.** \( \gamma \) is always strictly positive. Furthermore, when \( k \geq 2.87 \), the equilibrium pressure is \( \gamma = \left( (1 + 2k)^{\frac{1}{2}} - 1 \right) k^{-1}. \)

The absence of social pressure is never preferable to the agent. When perceived risk is low, the equilibrium \( \gamma \) is large but bounded. Boundedness follows because when \( k \) is small the equilibrium norm is large; if \( \gamma \) were large, the agent would then be working too hard for him to be able to take advantage from his concern for norms.

\[ \text{If the norm is contractible, then the contract consists of an incentive schedule together with a norm. The equilibrium contract is } \alpha = 0, \beta = \frac{1}{2}, a = \frac{1}{2}, \text{ and } \nu = \frac{1^2 + 2^2}{2k}, \text{ and the first best is attained. If such a contract was possible it would be preferable to the principal over any other arrangement.} \]

20
Proposition 6 allows us to perform asymptotic analysis.

Proposition 7. As $k \to \infty$, the equilibrium norm converges to the first best action. Furthermore, $\lim_{k \to \infty} a^N = \frac{1}{2}a^*$, $\lim_{k \to \infty} \alpha^N = 0$, $\lim_{k \to \infty} \beta^N = \frac{1}{8}$, $\lim_{k \to \infty} \nu = a^*$, $\lim_{k \to \infty} \frac{1}{2}k\gamma^2 = 1$, $\lim_{k \to \infty} U(\alpha^N, \beta^N, a^N) = \frac{1}{4}(U^* + V^*)$, and $\lim_{k \to \infty} V(\alpha^N, \beta^N, a^N) = \frac{1}{2}(U^* + V^*)$.

Note that total pressure $\left(\frac{1}{2}k\gamma^2\right)$ approaches 1 as perceived risk $k$ grows to infinity. The agent reduces $\gamma$ to zero as $k$ grows to infinity but not “fast” enough for pressure to be absent in the limit.

Therefore, even if $\gamma$ is endogenous, trade takes place even as the informativeness of the signal on effort vanishes. The limiting contract consists of a fixed payment and the agent takes an action that is half of the first best action.

3.5. Summary

It is always advantageous to the agent to develop norms against which to measure performance. Equilibrium norms are always high, but under sufficient risk they are not so high so as to leave the principal with no surplus. Because as perceived risk grows the equilibrium norm converges to the first best action and pressure to abide by the norm increases, moral hazard vanishes and trust increases. Most importantly, when pressure is exogenous, the outcome of the relationship approaches first best as the principal’s ability to provide incentives deteriorates. When pressure is endogenous, the limiting outcome as risk increases is $3/4$ of the first best outcome and surplus is shared between principal and agent. Finally, if the principal commits to a contract before the norm develops, the no-trust outcome takes place. This is preferable to the principal only in low-risk environments.

4. Ethical standards

By an ethical standard I mean a level of effort that the agent judges appropriate when he enters the agency relationship. To examine their role in agency, I consider acting self preferences

$$u = -\exp \left\{ -r \left( \alpha (a + x) + \beta - a^2 - \lambda (v - a)^2 \right) \right\}.$$  \hspace{1cm} (4.1)
These are just object self preferences (2.1) with the extra term

\[-\lambda (v - a)^2.\] (4.2)

The parameter \(v\) represents the ethical standard and \(\lambda\) represents the pressure felt by the agent to abide by the standard. This modeling approach was pioneered by Kandel and Lazear [13] who considered peer-pressure in deterministic partnership games.

The agent chooses \(\lambda\). When \(\lambda > 0\), the agent feels pressure to observe the standard because choosing an action other than the standard is costly (expression 4.2 is strictly negative). The further away is the action from the standard, the worse off the agent feels. Since the action is non-observable, the pressure captured by equation (4.2) is internal pressure and the agent’s distress is naturally interpreted as a feeling of guilt. When \(\lambda < 0\), the agent derives utility from dishonoring the ethical standard. Finally, when \(\lambda = 0\) the agent’s acting self preferences are a replica of his material payoffs, and just as in the standard model of agency, he does not care about ethical standards. Even though (4.2) is similar in form to the pressure function used to model norms (expression 3.2), the implications of ethical standards differ substantially from those of norms.

The timing of the game is as follows. In stage 1, before he meets with the principal the agent decides whether to develop a concern for ethical behavior \((\lambda > 0)\), not to modify his material self preferences \((\lambda = 0)\), or to become unethical \((\lambda < 0)\). In stage 2, principal and agent meet and the principal offers a contract to the agent \(h, \bar{\lambda}, a\) together with a speech in which she tells him what she considers to be ethical behavior \(\lambda\).14 The standard \(\lambda\) is thus chosen as part of the equilibrium by the principal. After the contract and the standard have been announced, the agent decides on whether to enter the relationship. If he does, then he chooses action \(a\).

The agent’s certainty equivalent of action \(a\) is given by

\[U^E = \alpha a + \beta - a^2 - \lambda (v - a)^2 - \frac{1}{2}k\alpha^2.\] (4.3)

14Such speeches are common. In the Kellogg Graduate School of Management, for example, every year the Dean addresses new professors and teaching Ph.D. students. He explains what is expected from them and what the school considers to be “ethical behavior.”
Note that when $\lambda > 0$, the agent is weakly worse off than if he did not take ethical standards into consideration because on top of the physical cost ($a^2$), he also feels the psychological cost of having to comply with the standard ($\lambda (v - a)^2$).

4.1. Rational ethics

The optimal action as a function of extrinsic incentives, the standard, and pressure is

$$a (\alpha (u (\lambda)), v (\lambda), \lambda) = \frac{\alpha + 2\lambda v}{2(1 + \lambda)}.$$  \hspace{1cm} (4.4)

Notice that larger extrinsic incentives and the larger ethical standard result in higher effort. The commission rate that maximizes total surplus given the ethical standard and the agent’s concern about the standard is

$$\alpha (u (\lambda), \lambda) = \frac{1}{2k(1 + \lambda) + 1}.$$  \hspace{1cm} (4.5)

The more pressured is the agent to comply with the ethical standard (larger $\lambda$), the less the extrinsic incentives given by the principal; there is a trade-off between intrinsic motivation and extrinsic incentives. The power of incentives, though, is independent of the standard, $v$. Intuitively, the principal wants the agent to choose an action as close as possible to first best. Given ethical standard $v$ and pressure $\lambda$, the first best action as evaluated by the principal is $\tilde{a}^* (v, \lambda) = \frac{1 + 2\lambda v}{2(1 + \lambda)}$. Substituting (4.5) in (4.4) we obtain the optimal action as a function of the ethical standard and pressure, $a (v, \lambda) = \frac{1 + 4\lambda v k + 4\lambda^2 v k + 2\lambda v}{2(2k + 2k\lambda + 1)(1 + \lambda)}$. To attain the first best the agent needs to be provided incentives to increase his effort by

$$\tilde{a}^* (v, \lambda) - a (v, \lambda) = \frac{k}{2k(1 + \lambda) + 1},$$

which is independent of $v$. On the other hand, since (4.6) is a decreasing function of $\lambda$, so is $\alpha$.

The ethical standard that maximizes total surplus given pressure $\lambda$ is

$$v (\lambda) = \frac{1}{2}.$$  \hspace{1cm} (4.6)
The first best action given the ethical standard \( v \) and pressure \( \lambda \) is 
\[
\hat{a}^* (v, \lambda) = \frac{1 + 2\lambda v}{2(1 + \lambda)}.
\]

We see that the principal chooses the ethical standard so that the first best action coincides with it. The equilibrium standard is independent of the level of pressure \( \lambda \) and of the level of perceived risk by the agent \( k \).

Because the agent’s concern about the ethical standard is genuine when he meets the principal, the fixed component of salary, \( \beta \), is set so that the agent is willing to participate given his ethical preferences. Making use of the normalizing assumption that the agent’s reservation certainty equivalent is zero, \( \beta \) is set so that
\[
\beta (\lambda) = -\alpha a + a^2 + \lambda (v - a)^2 + \frac{1}{2}ka^2.
\]

Finally, the agent chooses whether to become ethical or not. That is, the agent chooses \( \lambda \) by maximizing his material payoffs, subject to the principal’s participation constraint:
\[
\max_{\lambda} U \quad \text{subject to} \quad V \geq 0.
\]

The equilibrium level of self-imposed pressure is
\[
\lambda = \frac{2k + 1}{2k}, \tag{4.7}
\]

The principal’s participation constraint never binds and \( \lambda > 0 \) for all \( k \). Therefore, it is always in the best interest of the agent to develop ethical preferences and never to become unethical.

The equilibrium contract is
\[
\alpha^E = \frac{1}{2(1 + 2k)}, \quad \beta^E = \frac{k(k + 1)}{2(1 + 2k)^2}, \quad \text{and} \quad a^E = \frac{k + 1}{2(1 + 2k)}.
\]

The resulting certainty equivalents are
\[
U (\alpha^E, \beta^E, a^E) = \frac{\lambda k^2}{(1 + 2k + 2k\lambda)^2} \quad \text{and} \quad V (\alpha^E, \beta^E, a^E) = \frac{k + 1}{4(1 + 2k)}.
\]

I now plot the equilibrium values as functions of the level of perceived risk \( k \equiv r\sigma^2 \); see
Figure 3. The solid line corresponds the equilibrium of the ethical standards model. The dotted line portrays the equilibrium with material preferences.

The agent’s choice of pressure $\lambda$ can be thought of as the selection of a cost function. As noted above, the total cost of action $a$ consists of the physical cost ($a^2$) and the psychological cost ($\lambda \left( \frac{1}{2} - a \right)^2$). Graphically,

Figure 4: $a^2 + \lambda \left( \frac{1}{2} - a \right)^2$ as $\lambda$ increases
As pressure \( \lambda \) increases, \( a^2 + \lambda \left( \frac{1}{2} - a \right)^2 \) gets more convex and the action taken under no incentives gets closer to the standard. Since \( a^2 \) is always less than \( a^2 + \lambda \left( \frac{1}{2} - a \right)^2 \), an agent with a predisposition for ethical behavior, always bears more cost than one who does not consider ethical standards. This implies that \( \beta (\lambda) > \beta^S \). Also, since when \( \lambda > 0 \) the action chosen by the agent is closer to first best \( a^* \) for any given commission rate, we have that \( \alpha (\lambda) < \alpha^S \).\(^{15}\)

Computing the change in the incentive schedule as pressure increases, we see that \( \frac{d\alpha (\lambda)}{d\lambda} < 0 \) and \( \frac{d\beta (\lambda)}{d\lambda} > 0 \). The agent trades off one effect against the other. At each level of perceived risk \( k \), he keeps increasing \( \lambda \) up to the point in which the decrease in the commission rate \( \alpha \) results in a decrease of utility that is not compensated by the increased fixed payment \( \beta \).

Why is equilibrium pressure high when perceived risk is low? When \( k \) is low, the principal can induce the agent to take an action close to first best because risk bearing is low and she does not resent offering high powered incentives. No matter what the agent’s choice of \( \lambda \) is, he will end up choosing an action close to the first \( a^* = \frac{1}{2} \). By choosing \( \lambda \) large, the agent is at least able to reduce risk bearing.

The equilibrium has the following properties. First, \( \frac{da^E}{dk} < 0 \). Thus, just as in the standard model (section 2.2), the equilibrium action is reduced as risk grows. Nevertheless, \( a^E \) is bounded away from zero for all levels of perceived risk \( k \). Second, the equilibrium action is always less than the equilibrium ethical standard \( (v - a^E > 0) \). Finally, contrary to the standard model, the fixed component of wage \( \beta^E \) is always positive and principal and agent share total surplus.

Do ethical standards develop in high risk environments or they weaken as the level of perceived risk grows?

**Proposition 8.** \( \lim_{k \to \infty} \lambda = 1 \), \( \lim_{k \to \infty} \alpha^E = 0 \), \( \lim_{k \to \infty} \beta^E = \frac{1}{3} \), \( \lim_{k \to \infty} a^E = \frac{1}{2} a^* \), 
\[ \lim_{k \to \infty} U (\alpha^E, \beta^E, a^E) = \frac{1}{4} (U^* + V^*) \text{, and} \lim_{k \to \infty} V (\alpha^E, \beta^E, a^E) = \frac{1}{2} (U^* + V^*) \text{.} \]

\(^{15}\)One would think that in equilibrium the principal should choose an ethical standard that forces the agent to choose the first best action; the principal would then set \( \alpha = 0 \), and then compensate the agent through \( \beta \). This intuition misses that the first best action as evaluated by the principal is dependent on \( v \) and that the first best action is never chosen by the agent under no extrinsic incentives. Thus, the principal always sets \( \alpha > 0 \), no matter what the ethical standard is.
Thus, even if the pressure to abide by the standard weakens as \( k \) grows large, it is always larger than or equal to 1. This results in a limiting outcome equal to \( 3/4 \) of the first best total surplus, \( U^* + V^* \). Ethical standards, just as norms, rationalize Sen’s [23] observation that “no society would be viable without some norms and rules of conduct. Such norms and rules are necessary for viability exactly in fields where strictly economic incentives are absent and cannot be created.”

4.2. Trust

An agent who has developed pressure to abide by ethical standards, can be trusted to work up to \( a = \frac{2k + 1}{2(4k + 1)} \) without extrinsic incentives. It is easy to see that \( \alpha^E < \alpha^S, \beta^E > \beta^S \), and \( a^E > a^S \). Therefore, ethics-based trustworthiness results in the principal trusting the agent. Also, ethical standards result in larger total surplus than under untrustworthy preferences.

Finally, if the trust contract is offered and the agent acts to maximize his material payoffs, we see that the principal does worse than if she transacted directly with an untrustworthy agent. In particular, if \( k > \frac{1}{2} \), then she would have been better off by not entering into the relationship.

4.3. Committing to a contract

Would the principal ever wish to commit to a contract before the agent chooses whether to develop pressure to observe standards? As mentioned above, this amounts to changing the order of play in the game: the principal first offers a contract and then the agent chooses pressure and effort. The optimal pressure as a function of the incentive scheme is

\[
\lambda(\alpha, \beta) = 0.
\]

Thus, the agent prefers not to develop a concern over ethical standards. The reason is that with commitment the agent cannot affect the form of the equilibrium contract by his choice of pressure. The resulting equilibrium contract and outcome are as in section 2.2. Since the principal does better with ethical standards than without them, commitment is never
preferable to her.\textsuperscript{16}

4.4. Summary

It is always to the agent’s best advantage to develop pressure to abide by ethical standards and to the principal’s best interest to set ethical standards. The equilibrium standard as chosen by the principal corresponds to the first best action. Principal and agent share total surplus even if the principal has all bargaining power at the contracting stage. Contrary to the standard model, when the level of risk perceived by the agent goes to infinity, the equilibrium action converges to half of the first best action and total surplus does it to $\frac{3}{4}$ of the first best total surplus. An agent who has developed pressure to abide by ethical standards is trustworthy and is always trusted. Trust and trustworthiness result in larger surplus as compared to the no-trust situation. Finally, if the principal commits to a contract before the ethical standard develops, the no-trust outcome takes place.

4.5. The agent choosing the ethical standard

An interesting variation is having the agent choosing the ethical standard.\textsuperscript{17} I show that even if the principal has all the bargaining power at the contracting stage, the agent is able to appropriate all surplus. I begin by assuming that $\lambda$ is exogenous and positive and later I show that the agent will always set $0 < \lambda < \infty$. In stage 1, the agent chooses the standard $v$ and in stage 2 the agency relationship unfolds.

**Proposition 9.** In equilibrium, the principal’s participation constraint is binding. The equilibrium ethical standard is

$$v = \frac{1}{2} \left( 1 + \left( \frac{(1 + \lambda) (2k\lambda + 1)}{(1 + \lambda) 2k\lambda + \lambda} \right)^{\frac{1}{2}} \right).$$

\textsuperscript{16}If pressure $\lambda$ is contractible (observable and verifiable) so that the contract consists of an incentive schedule, an action, an ethical standard, and pressure to abide by the standard, then giving the principal full commitment power results in the first best outcome. The equilibrium contract is: $\alpha = 0$, $\beta = \frac{1}{2}$, $a = \frac{1}{2}$, $v = \frac{1 + \lambda}{2\lambda}$, and $\lambda > 0$. This would of course be preferable to the principal but the contractability of pressure is hard to justify.

\textsuperscript{17}This is a natural assumption when it is the agent who proposes the agency contract so that there is no announcement by the principal of what she considers to be ethical behavior.
Since the agent is compensated for the additional psychological cost through $\beta(v)$, he keeps increasing $v$ up to the point in which the principal is indifferent between contracting and staying out of the relationship. A simple computation shows that $\frac{dv}{dk} < 0$: as the level of perceived risk grows, the equilibrium standard is reduced. The reason is that as perceived risk $k$ increases, the total surplus that the agent can appropriate (through his choice of $v$) is reduced because risk bearing increases with $k$.

Assuming $\lambda = 1$, we can plot the equilibrium values as functions of the level of perceived risk ($k \equiv r\sigma^2$); see Figure 5. The solid line corresponds the equilibrium of the ethical standards model. The dotted line portrays the equilibrium with material preferences.

![Figure 5](image)

**Figure 5**: Effects of agent-chosen ethical standards on the equilibrium contract as the level of perceived risk, $k$, varies.

Just as with norms, the equilibrium ethical standard is larger than the first best action and the equilibrium action is less than the ethical standard. Furthermore, ethical standards take place in high risk environments. More precisely,

**Proposition 10.** $\lim_{k \to \infty} v = 1$, $\lim_{k \to \infty} a^E = \frac{\lambda}{1+\lambda}$, $\lim_{k \to \infty} \alpha^E = 0$, $\lim_{k \to \infty} \beta^E = \frac{\lambda}{1+\lambda}$.
\[
\lim_{k \to \infty} U(\alpha^E, \beta^E, a^E) = \frac{\lambda}{(1 + \lambda)^2}.
\]

It is immediate that when \( \lambda = 1 \), total surplus and the equilibrium action approach the first best as \( k \to \infty \). When \( 0 < \lambda \neq 1 \), first best does not take place but trade occurs. Thus, trade happens where orthodox principal-agent analyses predict a breakdown of the relationship.

It is not hard to see that an agent who develops ethical standards is trustworthy and the principal is willing to trust him. When the level of perceived risk is low, even though the principal trusts the agent and the agent honors trust, total surplus is less than if the agent was untrustworthy. When the level of perceived risk is large, trust results in larger total surplus.

Suppose that the principal commits to a contract before the agent chooses his ethical standard. The optimal action and ethical standard as a function of the incentive scheme are

\[
a(\alpha(v), v) = \frac{\alpha + 2\lambda v}{2(1 + \lambda)} \quad \text{and} \quad v(\alpha) = \frac{\alpha}{2}.
\]

Substituting \( v(\alpha) \) into \( a(\alpha(v), v) \), we see that \( a(\alpha) = v(\alpha) \) and thus no matter what the incentive schedule is, the agent chooses the ethical standard so that the equilibrium action coincides with it. This avoids psychological distress because now \( \lambda(v - a)^2 = 0 \) for all incentive schedules \( (\alpha, \beta) \). The equilibrium incentive schedule and total surplus can be shown to be exactly as when the principal faces a material agent directly (section 2.2). Thus, the no-trust outcome takes place. In light of proposition 9, I conclude that the principal would always like to commit to a contract and avoid the development of ethical standards.\(^\text{18}\)

An important question is whether it is in the agent’s best interest to develop ethical behavior; that is, whether it is in the agent’s best interest to have \( \lambda > 0 \). To analyze this, suppose there is a stage zero in which the agent chooses \( \lambda \). The following proposition shows that ethical behavior emerges endogenously.

**Proposition 11.** \( 0 < \lambda < \infty \).

\(^\text{18}\)If the ethical standard is contractible (observable and verifiable) so that the contract consists of an incentive schedule, an action, and an ethical standard, then giving the principal full commitment power results in the first best outcome. The equilibrium contract is: \( \alpha = 0, \beta = \frac{1}{4}, a = \frac{1}{2}, v = \frac{1}{2(1 + \lambda)} \). This would of course be preferable to the principal but the contractability of ethical standards is hard to justify.
To see this, note that since \( \lim_{\lambda \to 0} U(\alpha^E, \beta^E, a^E) = \frac{1}{4(2k+1)} \), the agent can always assure himself this much surplus. Also, since \( \lim_{\lambda \to 0^+} \frac{dU(\alpha^E, \beta^E, a^E)}{d\lambda} = \infty, \lambda = 0 \) is never optimal. Similarly, since \( \lim_{\lambda \to \infty} U(\alpha^E, \beta^E, a^E) = 0 \), very large \( \lambda \) is not optimal either. Since \( U(\alpha^E, \beta^E, a^E) \) is a continuous function of \( \lambda \), we conclude that the maximizing \( \lambda \) is strictly between 0 and \( \infty \).

Summarizing, it is always to the agent’s best advantage to develop ethical standards and also pressure to abide by those standards. Equilibrium standards are “high.” In fact, the equilibrium ethical standard is so high that the principal is left with no surplus. This is consistent with the high fixed fees charged by lawyers, consultants, and doctors; the standard model of agency is not. Also, if the principal can commit to a contract before the agent chooses the standard, then she is able lower the equilibrium ethical standard to a level that induces the agent to behave as in standard agency. Commitment is preferable to the principal because in the conventional model she ends up with positive surplus. Because equilibrium ethical standards may induce work above the first best action, they may also result in lower total surplus than if absent. Lastly, the equilibrium ethical standard is sufficiently high for the principal to be willing to trust the agent and as the level of perceived risk grows large, effort and total surplus are bounded away from zero even as extrinsic incentives vanish.

5. Altruism

By altruism I mean the agent’s willingness to act in the principal’s best interest originating from emotional identification or empathy.\(^{19}\) This suggests modeling altruism by incorporating a weighted version of the principal’s gross benefit function, \( \tilde{B}(a, x) = a + x \), in the agent’s utility. Gross benefit, as opposed to net benefit, is used for analytical tractability. It also has a natural interpretation: If one thinks of the principal as a representative shareholder and the gross benefit function as the random profit from the firm’s operations, then by caring about how the firm does, the agent (worker) is in fact caring about the principal’s

\(^{19}\)Adam Smith [24] was perhaps the first economist to point out that altruism is part of human nature: “How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it.”
well-being. Behavior payoffs are given by

\[ u = -\exp \{ -r (\alpha (a + x) + \beta - a^2 + \rho (a + x)) \} . \]

Parameter \( \rho \) reflects the level of altruism. A positive value is interpreted as altruism because the principal’s well-being positively affects the agent’s utility. Negative \( \rho \) is interpreted as spite because the agent feels worse the better off is the principal. The agent’s certainty equivalent of action \( a \) is

\[ U^A = \alpha a + \beta - a^2 + \rho a - \frac{1}{2} k (\alpha + \rho)^2 . \]  

(5.1)

Comparing (5.1) and (2.2) we see that altruism affects the agent’s certainty equivalent in two ways. First, given any \( \alpha \), he bears more risk because on top of an uncertain wage, he also feels the uncertainty embedded in the principal’s benefit function. A non-altruistic agent does not consider the variability in the principal’s gross benefit function and bears less risk. Second, abstracting from risk considerations, any given action is now less costly to the agent. In fact, the agent’s total cost function can be considered to be \( -a^2 + \rho a \). The first component, \( -a^2 \), corresponds to the physical cost of undertaking action \( a \). The other component, \( \rho a \), is the psychological pleasure derived from knowing that the principal has an expected gross benefit of \( a \). In summary, altruism increases risk bearing but reduces the cost of every action.

5.1. Rational altruism

I now examine the circumstances under which it is in the agent’s best interest to become altruistic. The optimal action as a function of extrinsic incentives and altruism is

\[ a (\alpha (\rho), \rho) = \frac{\alpha + \rho}{2} . \]  

(5.2)

Comparing this with the constraint in program (2.4) we see that for any given incentive schedule \( \langle \alpha, \beta \rangle \), an altruistic agent is always willing to work more than a non-altruistic agent. Consequently, altruism results in intrinsic motivation.
The principal optimally sets the commission rate by maximizing total surplus given the agent’s payoffs at the contracting stage. This yields
\[
\alpha (\rho) = \frac{1 - 2k\rho}{1 + 2k}. \tag{5.3}
\]

There is a trade-off between intrinsic motivation and extrinsic incentives: larger altruism results in lower extrinsic incentives. The fixed component of salary, \(\beta\), is set so that the agent is willing to participate given his payoffs at the contracting stage. That is,
\[
\beta (\rho) = -\alpha a + a^2 - \rho a + \frac{1}{2}k(\alpha + \rho)^2.
\]

The agent chooses altruism by maximizing his material payoffs subject to having the principal willing to enter the relationship. That is,

\[
\max_{\rho} U \quad \text{subject to} \quad V \geq 0. \tag{5.4}
\]

Equilibrium altruism is
\[
\rho = \frac{2k - 1}{2(1 + 2k^2 - k)}. \tag{5.5}
\]

Letting \(D = (1 + 2k^2 - k)(1 + 2k)\), we can write the equilibrium contract as
\[
\alpha^A = \frac{1}{D}, \quad \beta^A = \frac{(2k - 1)(1 + 4k^2)^2}{(4D)^2}, \quad \text{and} \quad a^A = \frac{(1 + 4k^2)}{4D}.
\]

In equilibrium the principal’s participation constraint is non-binding. Thus, total surplus is shared between principal and agent.

We can plot the equilibrium values as functions of the level of perceived risk \((k = r\sigma^2)\); see Figure 6. The solid line is the equilibrium in the model of altruism; the dotted line corresponds to the equilibrium with material preferences.
Figure 6: Effects of endogenous altruism on the equilibrium contract as the level of perceived risk, $k$, varies.

Note that altruism ($\rho > 0$) develops when $k > 1/2$ and spitefulness ($\rho < 0$) when $0 < k < 1/2$. Neither altruism nor spitefulness takes place when $k$ is either $1/2$ or $\infty$. The model delivers reasonable values. In particular, maximum altruism is $\rho = \frac{2\sqrt{2} - 1}{7} \approx 0.26$ and it occurs when $k = \frac{1 + \sqrt{2}}{2} \approx 1.2$. Maximum spitefulness is $\rho = -\frac{1}{2}$ and it happens at $k = 0$. Thus, the agent never regards the principal’s well-being as more important than his own.

As mentioned above, altruism results in larger risk bearing and in lower cost of each action. As a consequence, to implement any effort level, the principal can now offer lower powered incentives. Also, to persuade the agent to enter the relationship, she needs to offer a larger fixed wage to compensate for the additional risk bearing.

When the level of perceived risk is low ($k < 1/2$), altruism results in a decrease in the commission rate that is not compensated by a corresponding decrease in risk bearing. The agent is better off by becoming spiteful. This makes the principal increase the power of
incentives. The agent does not much resent the induced increase in risk bearing because his perception of risk $k$ is small. On the contrary, the increase in the commission rate is beneficial because now he earns a larger wage at any given effort level. In traditional production lines, jobs are well defined, there is direct monitoring, and the risk borne by the worker is low. The model suggests that spitefulness will develop and low surplus will be realized.

In medium to high risk environments, an increase in altruism results in a substantial increase in the risk borne by the acting self and in his willingness to work. Therefore, the principal reduces commission rate $\alpha$ to reduce risk bearing and increases $\beta$ to compensate for the lower bonus pay. The object self likes the reduction in risk bearing and the increase in the fixed pay, but he dislikes the lower total bonus pay. The agent trades off one effect against the other; he keeps increasing altruism as long as the reduction in risk bearing and the increase in the fixed payment compensates for the reduction in the monetary bonus.

When the level of perceived risk is very large ($k \to \infty$), the agent does not have much room to become altruistic. A small increase in $\rho$ results in a large increase in risk bearing and this makes the principal drastically reduce extrinsic incentives. This is undesirable to the object self. Hence, as $k \to \infty$, the willingness to become altruistic vanishes. Altruism is virtually non-existent when needed most, in those circumstances where the agent perceives so much risk that extrinsic incentives cannot be offered. In fact,

$$\lim_{k \to \infty} \rho = \lim_{k \to \infty} \alpha^A = \lim_{k \to \infty} a^A = \lim_{k \to \infty} U(\alpha^A, \beta^A, a^A) = \lim_{k \to \infty} V(\alpha^A, \beta^A, a^A) = 0.$$  

What drives the emergence of altruism in the present model is different from the case of partnership games studied by Rotemberg [21]. In a partnership, agents become altruistic when their choices are strategic complements. Becoming altruistic increases the acting self willingness to work and this, by complementarity, makes other agents’ effort increase also. In the principal-agent model, the agent becomes altruistic to affect the equilibrium contract to his best advantage.
5.2. Trust

I now show that altruism fosters the emergence of trust. Without extrinsic incentives the agent chooses action
\[ a = \rho a^*, \]
which is positive when \( \rho > 0 \). We see that an altruistic agent can be trusted to work even in the absence of monetary incentives. In fact, he works a proportion \( \rho \) of the first best action, \( a^* \). Therefore, altruistic preferences are trustworthy preferences according to Definition 1.

One can show that \( \alpha^A \) is less (greater) than \( \alpha^S \) when \( \rho > (\leq) 0 \) and \( a^A \) is greater (less) than \( a^S \) when \( \rho > (\leq) 0 \). Also, \( \beta^A \) is always larger than \( \beta^S \). Therefore, when the agent is altruistic, the principal trusts him. Unfortunately, since \( \rho \to 0 \) as \( k \to \infty \), trust is virtually non-existent when risk is large. Also, the principal’s certainty equivalent is larger than under untrustworthy preferences only when the agent develops altruism, that is, only if \( k > \frac{1}{2} \). Therefore, I can state

**Proposition 12.** If the agent is altruistic, then he is also trustworthy, the principal trusts him, and superior gains are realized. A spiteful agent is never trusted.

Finally, suppose that the agent is offered the altruism contract. Let \( V' \) be the surplus to the principal if the agent acts to maximize his material payoffs. It is easy to see that \( V(\alpha^A, \beta^A, a^A) > V' \) only when \( k > \frac{1}{2} \). The principal ends up worse off only if she has trusted the agent and the agent has acted to maximize his material payoffs. In fact, when \( k > \frac{1 + \sqrt{2}}{2} \), then \( V' < 0 \), and she would have been better off by staying out of the relationship (rather than offering the trust contract) if the agent cared only about material payoffs.

5.3. Committing to a contract

If altruism is non-contractible (it is observable but not verifiable), then we see that whatever the contract \( \langle \alpha, \beta, a \rangle \) is, the agent will set \( \rho (\alpha, \beta) = 0 \). Thus, commitment prevents the development of altruism and consequently that of trust. The reason is that now the agent’s choice of altruism cannot affect the shape of the equilibrium incentive scheme. The optimal commitment contract is \( \langle \alpha^S, \beta^S, a^S \rangle \) as in section 2.2. In light of the results in the sections
above, the principal prefers commitment when \( k < \frac{1}{2} \) so that she avoids the emergence of spite.\(^{20}\)

The conventional model of the principal-agent relationship, portrayed in section 2.2, implicitly assumes that the agent cannot credibly modify his acting self preferences, even though such an opportunistic change in tastes may be advantageous to both principal and agent.

### 5.4. Deterministic benefit function

It is interesting to study the important special case in which the principal’s gross benefit function \( \tilde{B} \) is deterministic. Suppose \( B = a \). Then, equilibrium altruism is

\[
\rho = \frac{-1}{2 + 4k} < 0.
\]

The agent prefers to be spiteful, for all levels of perceived risk \( k \). When the principal’s gross benefit function is deterministic, altruism does not result in larger risk bearing; it only results in lower cost for any action as compared to the standard model of agency. Thus, if the agent becomes altruistic, then the principal offers him lower fixed payment and the same commission rate as in the conventional model. This is undesirable to the object self. In contrast, spite (\( \rho < 0 \)) results in a larger fixed payment because now the agent bears larger cost at every effort level.\(^{21}\)

In the conventional model of agency (section 2.2), the principal is indifferent between benefit function \( B(a) = a \) and \( \tilde{B}(a, x) = a + x \) because she is risk neutral and \( E\tilde{B}(a, x) = \)

---

\(^{20}\)If altruism is contractible (observable and verifiable) so that the contract consists of an incentive schedule, an action, and a level of altruism, then giving the principal full commitment power results in the first best outcome. The equilibrium contract is: \( \alpha = 0, \beta = \frac{1}{4}, a = \frac{1}{2}, \rho = 1 \). Thus, the first best level of altruism is having the agent caring about the principal as much as he cares about himself. Lower and, perhaps more surprisingly, larger altruism is total welfare reducing. Even though this contract is preferable to the principal, the assumption that altruism is contractible is hard to justify.

\(^{21}\)I have generalized this model to the case in which the principal’s benefit function depends on the realization of a normally distributed random variable with mean \( a \) and variance \( \xi^2 \). I allow the random variables \( a + x \) and \( a + y \) to be jointly distributed (bivariate normal) with correlation coefficient \( \kappa \). (Because of bivariate normal, the covariance is \( \kappa \sigma \xi \); also, \( a + x \) and \( a + y \) are independent if and only if the covariance is zero.) Two points are worth mentioning. First, when \( a + x \) and \( a + y \) are uncorrelated (and thus are independent), there are no incentives for altruism (or spite) to arise. Second, when \( \xi^2 \) is very large relative to \( \sigma^2 \), the optimal level of altruism is \( \rho = \infty \).
When altruism is a possibility, the principal may be strictly better off with a gross benefit function that has randomness even if she is risk neutral and \( E\tilde{B}(a,x) = B(a) \) because such randomness allows altruism to develop.\(^{22}\)

### 5.5. Summary

Agency fosters the emergence of altruism. Maximum altruism arises in medium risk environments. High risk also promotes altruism, but as the level of perceived risk grows large, the coefficient of altruism tends to zero. An altruistic agent is always given lower powered incentives and a larger fixed salary than if he was non-altruistic; he is also intrinsically motivated to work. Altruism results in the principal trusting the agent. Those environments requiring most trust do not foster the emergence of altruism; thus, the role of altruism in promoting trust in agency is limited. Also, for altruism to arise, the principal’s gross benefit function needs to have randomness. Low perceived risk promotes spitefulness, distrust, and inferior outcomes. Finally, the conventional model can be interpreted as one where the principal commits to a contract before the agent chooses whether to develop altruism.

### 6. Comparison

I have studied norms, ethical standards, and altruism independently. I now summarize the relative advantages of each of the motivational schemes to both principal and agent.

For low values of \( k \), ethical standards are best for the principal. When the agent’s perception of risk is large, norms induce the largest surplus to the principal. Thus, in those situations where extrinsic incentives cannot be given because the agent perceives large risk, norms are the preferable intrinsic motivational scheme to the principal. Transacting with an altruistic agent results in more surplus than if the agent is non-altruistic, but to the principal, altruism is always dominated by ethical standards.

Under untrustworthy preferences, the agent is left with no surplus. Norms are more advantageous to the agent than ethical standards for small values of perceived risk \( k \). As \( k \) grows large, the agent’s surplus under norms goes to zero, whereas the surplus under ethical

\(^{22}\)\( B(a,x) = a + x \) is preferable to \( B(a) = a \) as long as \( t > \frac{1}{\epsilon} \).
standards is bounded away from zero. If parameter $\gamma$ in the model of norms is large, then altruism may be advantageous to the agent for medium values of perceived risk $k$.

7. Conclusion

Agency contracts depend fundamentally on trust when it is difficult to provide extrinsic incentives. Extrinsic incentives may be silent because the agent is highly risk averse, there is no informative signal on the agent’s choices, or the agent allocates a homogeneous input among diverse activities. Without trust, the agency relationship may break down.

I have examined three reasons for the emergence of trust in agency: norms, ethical standards, and altruism. Norms are strongest in promoting trust because they converge to the first best action as the principal’s ability to provide extrinsic incentives deteriorates. Ethical standards lead to $3/4$ of the first best outcome. Altruism plays a minor role in promoting trust because it is nearly absent in high risk environments.

The results in all three models are consistent with organizational arrangements that increase worker’s awareness to risk. If the worker’s job is designed so that he is exposed to large risk, mutually beneficial norms, ethical standards, and altruism will develop and better outcomes will be obtained. This may be one reason for the increasing presence of organizational arrangements such as participative management. To the extent that the worker’s perception of risk is increased, these instruments will result in larger intrinsic motivation as well as in the worker’s identification with the firm’s objectives.

In contrast to the standard agency model, agency with trust results in total surplus being shared between principal and agent. Moreover, the first best outcome may ensue in highly uncertain environments. Also, the principal may be better off the more the agent is risk averse. Further, it is easy to construct examples for each of the motivational schemes in which larger extrinsic incentives (due to a lower perception of risk) are in equilibrium associated with lower actions and/or lower total surplus. Thus, contrary to the orthodox model, the motivational schemes I have considered are consistent with Kreps’s [15] observation that larger monetary incentives are often associated with lower effort and lower total surplus because intrinsic motivation is also lower.
The results can be extended to Holmstrom and Milgrom’s [12] multitasking environment. The agent chooses not only how much effort to exert, but also how to allocate effort among different activities. Providing higher incentives on any one task induces the agent to devote all his effort to that task, and none to the rest. As a consequence, incentives need be the same for all tasks. If a task that is essential to the principal has a very noisy information signal, incentives must be reduced for all tasks. As the variance of the signal on the effort exerted on this essential task goes to infinity, overall incentives must be eliminated. Yet, why should the agent be expected to work if no incentives can be provided? Elsewhere, I show that an effort-averse agent will develop ethical standards and norms that will allow for trade to occur. Holmstrom and Milgrom’s [12] no-incentives result can be interpreted as a theory of contractual incompleteness: principal-agent contracts are silent on many aspects of the agent’s performance because otherwise overall incentives would be distorted and this would result in inferior outcomes. With this interpretation, trust becomes necessary for the functioning of incomplete contracts.

Concepts and ideas from social psychology and sociology have proven very fruitful in enhancing our understanding of the principal-agent relationship. Intrinsic motivational schemes such as norms, ethical standards, and altruism together with trust help explain the form of observed agency contracts. Accordingly, it is necessary to incorporate these notions into the standard model of the principal-agent relationship.
8. Appendix

Proof of proposition 1: The unconstrained optimum is

$$\nu = \begin{cases} 
\infty & \text{if } 2k - 1 - k\gamma + k^2\gamma^2 \leq 0 \\
\frac{k(2 + k\gamma)^2(1 + k\gamma) + (2 - \gamma)(1 + k\gamma)^2}{2\gamma(k^2(2 + k\gamma)^2 - (1 + k\gamma)^2)} & \text{if } 2k - 1 - k\gamma + k^2\gamma^2 > 0.
\end{cases} \quad (8.1)$$

If $\nu$ is given by (8.1), then the principal’s participation constraint is violated when $2k - 1 - k\gamma + k^2\gamma^2 \leq 0$. When $2k - 1 - k\gamma + k^2\gamma^2 > 0$, the principal’s participation constraint under (8.1) becomes

$$(1 + k\gamma) \frac{1 + 3k\gamma + 8k^4\gamma + 12k^5\gamma^2 + 6k^6\gamma^5 + 3k^2\gamma^2 + k^7\gamma^7 + k^3\gamma^3 - k^2(2k\gamma + 2k + 2)(2 + k\gamma)^2}{4(k^2\gamma^2 + 1 + k\gamma + 2k)^2(k^2\gamma^2 - 1 - k\gamma + 2k)} \geq 0. \quad (8.2)$$

For fixed value of $\gamma$, this inequality is strict if $k$ is large enough. Fix $\gamma$ and let $k(\gamma)$ be the largest $k$ such that (8.2) is satisfied with equality. Since for all $k$ such that $2k - 1 - k\gamma + k^2\gamma^2 > 0$, $\frac{dV}{dk} > 0$, we have that for $k > k(\gamma)$,

$$\nu = \frac{1}{2} \frac{k(2 + k\gamma)^2(1 + k\gamma) + (2 - \gamma)(1 + k\gamma)^2}{\gamma(k^2(2 + k\gamma)^2 - (1 + k\gamma)^2)}$$

is the equilibrium norm. When $k \leq k(\gamma)$, the equilibrium norm results in a binding principal’s participation constraint. There are two values of $\nu$ such that $V = 0$,

$$\nu_2 = \frac{1}{2} \left(1 - \frac{1}{\gamma} \left(\left(\frac{(2k+1)(2+k\gamma)}{k}\right)^{\frac{1}{2}} + 2\right)\right) \quad \text{and} \quad \nu_1 = \frac{1}{2} \left(1 + \frac{1}{\gamma} \left(\left(\frac{(2k+1)(2+k\gamma)}{k}\right)^{\frac{1}{2}} - 2\right)\right)$$

Substituting in $U$, we see that $U(\nu_1) > U(\nu_2)$. ■

Proof of proposition 2: The first best action is $a^* = \frac{1}{2}$. Now,

$$a^* - \nu = \begin{cases} 
-k\gamma + \sqrt{4k^2 + 2k^3\gamma^2 + 2k + k^2\gamma^2} & \text{if } k \leq k(\gamma) \\
\frac{k^3\gamma + 6k^2\gamma^2 + 4k\gamma + 4k + 2}{2\gamma(2k - 1 - k\gamma + k^2\gamma^2)(2k + 1 + k\gamma + k^2\gamma^2)} & \text{if } k > k(\gamma).
\end{cases} \quad (8.3)$$
Also

\[ a^N - \nu = \begin{cases} 
- \frac{k\gamma + 2k + k^2\gamma^2}{\gamma((2k+1)(2+k\gamma^2)k)^2} & \text{if } k \leq k(\gamma) \\
- \frac{1}{(2k-1-k\gamma+k^2\gamma^2)} & \text{if } k > k(\gamma).
\end{cases} \]  

(8.4)

Both (8.3) and (8.4) are negative because for all \( k > k(\gamma), 2k - 1 - k\gamma + k^2\gamma^2 > 0. \)

**Proof of proposition 4:** I need only show that when \( k \) is large, \( a^N < a^S \) because proposition 3 guarantees that \( \beta^N > \beta^S \) and \( a^N > a^S \) for large \( k \). When \( k > k(\lambda) \),

\[
\alpha^S - \alpha^N = k \frac{4k + 6k^2\lambda^2 + k^34 + 4k\lambda + 2c}{(2k - 1 - k\lambda + k^2\lambda^2)(2k + 1 + k\lambda + k^2\lambda^2)(1 + 2k)} > 0
\]

because when \( k > k(\lambda), 2k - 1 - k\lambda + k^2\lambda^2 > 0. \)

**Proof for proposition 5:** Suppose that the norms contract is offered but the agent acts to maximize his material payoffs. Suppose first that \( k \leq k(\gamma) \). Then,

\[
V = -\frac{1}{2} \frac{(1 + k\gamma) \gamma k}{(2k + 1)} < 0.
\]

When \( k > k(\gamma) \), we have that

\[
V = -\frac{1}{4} (1 + k\gamma)^2 \frac{2(2k+k^2\gamma^2)^2(1+k\gamma)^2+(2k+1)(1+k\gamma)^2+6k^4\gamma^2+6k^6\gamma^6}{(4k^2+4k^3\gamma^2+k^4\gamma^4-2k\gamma-1-k^2\gamma^2)^2}.
\]

But since in this zone \( 2k - 1 - k\gamma + k^2\gamma^2 > 0 \), we have that

\[
2k - 1 - k\gamma + k^2\gamma^2 > 0 \\
\Rightarrow 2k + k^2\gamma^2 > 1 + k\gamma \\
\Rightarrow (2k + k^2\gamma^2)^2 > (1 + k\gamma)^2.
\]

Therefore, \( V < -\frac{1}{4} (1 + k\gamma)^2 \frac{8k^3+12k^4\gamma^2+6k^6\gamma^4+k^6\gamma^6}{(4k^2+4k^3\gamma^2+k^4\gamma^4-2k\gamma-1-k^2\gamma^2)^2} < 0. \)

**Proof for proposition 6:** When \( \nu = \frac{1}{2} \frac{k(2+k\gamma)^2(1+k\gamma)^2+(2(1+k\gamma)^2)}{(k^2(2+k\gamma^2)^2-(1+k\gamma)^2)} \), the agent’s certainty equivalent is \( U = \frac{k(k+1)^2}{2(2k-1-k\gamma+k^2)(2k+1+k\gamma+k^2)} \). This is maximized at \( \gamma = \left( \frac{(1 + 2k)^{\frac{2}{3}} - 1}{k} \right)^{k^{-1}}. \)

At \( \gamma = \left( (1 + 2k)^{\frac{2}{3}} - 1 \right)^{k^{-1}} \) we have that \( V \) is larger than zero as long as \( k > 2.87 \). When
\( k < 2.87, \quad \frac{d(\alpha a + \beta - \alpha^2 - \frac{1}{2}k\alpha^2)}{d\gamma} \bigg|_{\gamma=0} = \frac{1}{2} \frac{k}{(1 + 2k)} > 0. \) Therefore, \( \gamma > 0. \) ■
References


