How to build stable relationships between people who lie and cheat.

Ivar Ekeland
CEREMADE, Université Paris-Dauphine

March 2013

Abstract

This is a talk delivered at the conference "Mathematics in a Complex World", on the occasion of the 150th anniversary of the Politecnico di Milano. Asymmetry of information, i.e. the possibility for human beings to hide their information, or not to keep their promises, is a fundamental fact of social life, and must be taken into account. I will show how this creates complexity, even in the very simple situation of a contract between two parties, one of whom commits to work for the other, but cannot be monitored.

1 Building society from the bottom up

There are two approaches to the social sciences. One can go bottom-up, starting from free individuals and trying to understand how they form societies, or one can go top-down, starting from a given society and trying to understand how they assimilate individuals. The first one goes back to the bourgeois Enlightenment, marked by Rousseau and Montesquieu in France, Locke in England, Hume and Smith in Scotland, while the second one originates with Comte, Hegel and Marx, and has given birth to sociology and anthropology. One can also seek analogies in the sciences, and relate the first approach to physics, and the second to biology. A physicist will consider atoms and molecules as the primary entities, and will seek to explain the properties of macroscopic objects, solids, liquids or gases, from the properties of these microscopic building blocks. A biologist, on the contrary, will consider the macroscopic organism as the primary entity, and will try to understand how its microscopic components fit into the bigger entity and help it fulfil its purpose.

Modern economic theory is decidedly bottom-up: it considers that societies are constructed from individuals, much like are built from atoms. It then has to explain (a) how isolated individuals behave, and (b) how they interact with others.

Individuals are characterized by a set of preferences, which determines their behaviour: in each circumstance, they are supposed to rank all possible al-

ternatives, and to choose the one they prefer. This is usually expressed by a utility function u over possible actions. If individual 1 has a utility function u_1, and $u(a) > u(a')$, then she prefers a to a'; note that another individual will have another utility function u_2, and we may well have $u_2(a) < u_2(a')$, so that the two individuals have opposite tastes. Of course, the consequences of any decision are a usually not known with total certainty: there are several possible outcomes x, each of which has a probability P^a, which depends on a, and the agent then ranks the outcomes, not the decision itself. She is supposed to choose the action which maximises the expected utility of the outcomes, that is, to choose a so as to maximise $\mathbb{E}^a[u(x)]$. The function u is concave, which captures various degrees of risk aversion. If we are dealing with money, so that x is one-dimensional, one usually takes $u(x) = \frac{1}{\alpha}x^\alpha$, with $\alpha \leq 1$ (negative values of α are admissible, and for $\alpha = 0$ one takes $u(x) = \ln x$), with $x \geq 0$. If $\alpha = 1$, the corresponding individual is risk-neutral: she will pay $50,000$ to buy a lottery ticket which gives her a 50% chance of winning $100,000$. Most individuals would pay much less for such a ticket; how much they would pay depends on their particular α, and the amount is easily seen to decrease with α.

These individuals, each one with his or her utility function, are now supposed to enter contracts with each other. There is a large variety of possible contracts. At one end of the scale, there is the simple act of buying bread: you give a loaf or a baguette, and I give you cash. At the other end, you have the constitutions and laws regulating states, which in fact are contract between its citizens. There is even a long-standing tradition in political science, going back to Rousseau, which imagines the free individuals, in the so-called " state of nature" (meaning that society does not exist yet), gathering together before history begins and negociating their freedom away in the "social contract", behind a suitable "veil of ignorance" (Rawls). In short, one understands society as a network of contracts between individuals and groups. This is obviously a highly complicated view of society, much more complicated then, say, the physicist's view of solids: atoms come in finitely many species, and within each species they are identical. They do not have memories, they are not born, they do not age and they do not die. Human beings, on the other hand, are all different, and they change during their lifetime. We are very far from being able to understand how society emerges from this variety, so we will be content with studying the simplest possible contract, which is a contract between two parties.

The most important fact that the economist has to contend with, is assymmetry of information. If human beings were transparent to each other, social life would be easy. One would detect liars because their nose grows, like Pinocchio's, and one would detect evil people because they have black tails and hooved feet. Unfortunately it does not work that way. People can lie, or at least hide relevant information, and people can renege on promises, as every voter will tell you. This is called asymmetry of information, and economists classify it in two categories: adverse selection and moral hazard.

Adverse selection consists in hiding information or outright lying. I know

things about myself that I am not going to tell you. If I am interested in buying your house, I will not rave about how wonderful it is and how much I would be willing to spend to get it. It is wiser to let the seller think that I am interested, but that I am looking at other houses, and anyway I do not have deep pockets. Adverse selection takes its name from the insurance industry. In health insurance, for instance, the insurer would like to know the genetic imprint and past history of all applicants, but this would result in insuring only those who do not need it, namely the healthy !

In the short span of this lecture, I will not deal with adverse selection, and illustrate instead the contractual consequences of moral hazard. As I mentioned, people will renege on promises, or simply cheat, if they can get aay with it. When drawing up a contract, this possibility has to be taken into account, and this results in great complications. In fact, these complications where insuperable until very recently, when Sannikov found a way to reduce the moral hazard problem to an optimal control problem in a seminal paper of 2008 [5]. Since then, his method has been shown to be extremely fruitful (see [6]; we will give more references later on). In this conference, we will describe Sannikov's method. As usual, there are two approaches to the resulting optimal control problem, one through PDEs and the HJB (Hamilton-Jacobi-Bellman) equation, and the other through the stochastic maximum principle and BSDEs (backward stochastic differential equations). We will choose the first one, like Sannikov, and we refer to the recent book of Cvitanic [3] for the other.

2   Working for someone else

We will descrive a simple contract between two parties, where the first one (the principal) hires the second one (the agent) to work for him. The agent is in charge of a project which generates a stream of revenue dX_t for the principal

$$dX_t = A_t dt + \sigma dZ_t$$

Here Z_t is a standard Brownian motion, which generates a filtration $\mathcal{F}_t^Z$, and $\sigma > 0$ is a given constant. The process A_t is $\mathcal{F}_t^Z$-adapted, with $0 \leq A_t \leq \bar{a}$. It is chosen, not by the principal, but by the agent. If the agent chooses A_t at time t, she incurs an instateneous cost of $h(A_t)$, where $h : [0, \bar{a}] \to \mathbb{R}$ is an increasing and convex function, with $h(0) = 0$. We shall refer to A_t as the effort of the agent, and to $h(A_t)$ as the cost of effort. Since the revenue dX_t accrues to the prinicipal, the agent has no interest in making any effort, and would choose $A_t = 0$. So the principal has to compensate her for effort, and to this end chooses a process $C_t \geq 0$. Let us now write the intertemporal utilities of both parties:

$$\text{(principal) } r\mathbb{E}\left[\int_0^\infty e^{-rt}(dX_t - C_t dt)\right] \tag{1}$$

$$\text{(agent) } \quad r\mathbb{E}\left[\int_0^\infty e^{-rt}(u(C_t) - h(A_t)) dt\right] \tag{2}$$

As mentioned earlier, A_t is the agent's effort (chosen by the agent) and C_t is the agent's compensation (chosen by the principal). Both parties have the same discount rate r (the presence of r in front of the integrals has no role except to somewhat simplify the formulas). The remaining function u and h characterize the agent: u is her utility function, and h the cost of effort. Note that the agent is risk-adverse but not the principal.

If we made no further assumptions, the principal would want the agent to work as hard as possible, so that $A_t = \bar{a}$, and would propose her a contract whereby she sets $A_t = \bar{a}$ and gets in exchange a compensation slightly above the level $\bar{c}$ wheret $u(\bar{c}) = h(\bar{a})$; the agent would then accept the contract, because her intertemporal utility then is slightly above 0. In other words, the agent would be kept at subsistence level while the principal would capture all the surplus. We now incorporate in the model two major features, which will put the agent in a much better (in fact, as we shall see, a fantastically better) position: moral hazard and limited liability.

Condition 1 (1:Moral Hazard) *Denote by $\mathcal{F}_t^X$ the filtration generated by X_t. Then the process C_t is $\mathcal{F}_t^X$-adapted*

In other words, the agent's compensation is conditional on past revenues: C_t can depend only on the values of X_s for $s \leq t$. This expresses moral hazard: the principal observes neither A_t nor Z_t, he observes only dX_t. The randomness protects the agent: if the principal's revenue decreases, the agent can always claim that it is due to bad luck ($dZ_t < 0$), rather than lack of effort ($A_t < 0$). The agent would dearly like to make compensation conditional on present effort, that is, set $C_t = \varphi(A_t)$ for some increasing function φ, but he cannot, because the agent can always shirk, i.e. claim to have done maximum effort ($A_t = \bar{a}$), while in fact doing nothing ($A_t = 0$). The principal can only make the contract conditional on what he can observe, namely X_t. Of course, the past history of X_t, that is, the values of X_s for $s \leq t$, contain information about the agent's actual performance, and the right contract should be able to extract that information. So it is certainly not going to be Markovian, i.e. C_t should not depend on t only, but on all the history of X_s for $s \leq t$.

Condition 2 (2: Limited Liability) $C_t \geq 0$ *for all t.*

This means that the principal cannot fine the agent, or even recover money given earlier: any compensation given to the agent is for ever lost to the principal. The principal would dearly like the contract to specify some dreadful penalty if the agent is caught napping on the job; if the penalty is severe enough, the agent is more or less sure to perform, even if the likelihood of being caught is small. Fortunately (for the agent) he cannot do that, barring exceptional circumstances (sentries in time of war are supposed to be shot if they fall asleep): the worst that can happen to her is to be fired. Since the principal does not have a stick, he will have to use the carrot: instead of punishing severely poor performance, he will have to reward extravagantly good performance.

Before we proceed with the mathematical analysis, let us give some examples. Suppose I am an absentee landlord: I own a farm, in some remote countryside far away from the big city where I live, and I want someone else to take care of it for me. What kind of contract should I give the tenant ? Three possibilities come to mind:

1. $C_t = c$ (fixed salary). The agent gets c no matter what the revenue is. There is no incentive for her to work, she gets c whether $A_t = 0$ or $A_t = \bar{a}$, and since effort is costly she chooses $A_t = 0$.

2. $C_t = aX_t$ (fixed share of revenue). The agent gets to share the revenue. She now has an incentive to work - the harder she works, the higher is her expected gain. The problem for the principal is to choose the correct value of a, that is, not to be too generous: the higher a, the higher the agent's incentive, but the higher the cost to the principal, all the way up to $a = 1$ where all the revenue goes to the agent

3. $C_t = X_t - c$ (the principal rents out the project to the tenant). This seems an excellent compromise: the agent's incentive is maximal, and yet the principal gets something as well. The problem, of course, is to determine the correct value of c. There is another drawback to this solution: it is not an efficient way to share the risk. The principal is risk-neutral, whereas the agent is risk-adverse, so the principal should bear the risk. And yet, with $C_t = X_t - c$, it is the agent who bears all the risk, in fact insuring the principal

The situation I have described occurs in all kinds of societies, where absentee landlords, living in cities, have tenants cultivate their dominions. Solution 1 (fixed salary) is never observed. Note, however, that is used in other situations, such as domestic help or farmhands, where the agent works under the close supervision of the principal, so that there is little or no moral hazard. Solution 2 is known as sharecropping, and Solution 2 as farming. Both are historically and geographically widespread, and there are very interesting economic studies characterizing conditions favouring one or the other.

The same situation occurs in the modern finance industry. Suppose you have $100,000$ Euros to invest - say, savings towards your retirement. You are not going to invest them yourself: you have neither the time nor the expertise. You bring them to an investment manager, who will invest them for you. And then you are back into the moral hazard problem: if the returns are bad, is it due to her or to the market ? It is easy to blame everything on the economic environment, and you are very unlikely to hear any other explanation, but how do you know ? Even when bad results persist, and you become convinced that your manager is at fault, that she is regularly taking too much risk and picking the bad investment options, your only resort is to fire her, that is, to cash in whatever remains of your portfolio and put it away. All the losses are yours, and there is no recovering the management fees: this is the limited liability problem.

Moral hazard and limited liability are the core of the finance industry. The money manager invests other people's money and charges a fee for doing so. Financial investments are so arcane that the client cannot monitor what the manager is actually doing (moral hazard), and since the client, not the manager (limited liability) is carrying the losses, the manager is prone to taking risks which he would not take with his own money.

We summarize the preceding discussion:

Definition 3 *A* contract *is a pair* (C_t, A_t), *with* $C_t \geq 0$ *and* $\mathcal{F}_t^X$*-adapted while* A_t *is* $\mathcal{F}_t^Z$*-adapted*

The point to bear in mind is that the contract is written down at time $t = 0$. The compensation part C_t is verifiable (both parties observe X_t) and enforceable (the principal has to pay C_t; if he does not, the agent goes to court and forces him to do so). Note that this precludes any kind of arbitrary behaviour. The principal can specify in the contract under which circumstances he would fire the agent, and he will fire him if these circumstances are met; but he cannot fire the agent under other circumstances. The effort part A_t is specified in the contract, but cannot be verified, since only the agent observes it.

On paper, the contract binds both parties for ever. In effect, it can be terminated by either party:

- by setting $C_t = 0$ for $t \geq T$ the principal fires the agent at time T

- by setting $A_t = 0$ for $t \geq T$ the agent walks off the job at time T

Definition 4 *A contract* (C_t, A_t) *is* incentive-compatible *(IC) if the agent finds it in her own interest to exert the contractual effort* A_t *at every t. It is* individually rational *(IR) if both the principal and the agent find it in their own interest to enter the contract at* $t = 0$.

$$\text{(principal) } r\mathbb{E}\left[\int_0^\infty e^{-rt}\left(dX_t - C_t dt\right)\right] \geq 0 \tag{3}$$

$$\text{(agent) } \quad r\mathbb{E}\left[\int_0^\infty e^{-rt}\left(u\left(C_t\right) - h\left(A_t\right)\right) dt\right] \geq 0 \tag{4}$$

The principal devises a contract and offers it to the agent. The agent can take it or leave it: there is no negociation. If a contract does not satisfy (3), the principal will not offer it, for he is better off doing nothing. If a contract does not satisfy (4), the agent will not accept it, for she is better off doing nothing. If a contract is not IC, the principal will not offer it, for he knows the agent while she will not deliver her side of the bargain. The *principal's problem* consists of maximizing his expected profit (1) among all contracts which are IR and IC.

3 Solving the principal's problem.

3.1 Characterizing IC contracts

Suppose a contract (C_t, A_t) is given. Let us look at the *continuation value* the agent derives from it at time t:

$$
\begin{aligned}
W_t &= r\mathbb{E}\left[\int_t^\infty e^{-r(s-t)}\left(u\left(C_s\right) - h\left(A_s\right)\right) ds \mid \mathcal{F}_t^Z\right] \\
&= re^{rt}\mathbb{E}\left[\int_0^\infty e^{-rs}\left(u\left(C_s\right) - h\left(A_s\right)\right) ds \mid \mathcal{F}_t^Z\right] - re^{rt}\int_0^t e^{-rs}\left(u\left(C_s\right) - h\left(A_s\right)\right) ds
\end{aligned}
$$

The first term on the RHS is a martingale. Using the martingale representation theorem we find that there is a Z_t-adapted process Y_t (depending on C_t and A_t) such that:

$$
\begin{aligned}
\frac{1}{r}dW_t &= \left(W_t - u\left(C_t\right) + h\left(A_t\right)\right) dt + Y_t\sigma dZ_t &&(5) \\
&= \left(W_t - u\left(C_t\right) + h\left(A_t\right) - Y_t A_t\right) dt + Y_t dX_t &&(6)
\end{aligned}
$$

Now suppose the agent has conformed to the contract (C_s, A_s) for $s \leq t$, and tries to cheat, by performing effort a in the following interval $[t,\ t+dt]$, and reverting to A_s for $s \geq t+dt$:

- her immediate compensation C_t is unaffected

- her cost on $[t,\ t+dt]$ is $rh\left(a\right) dt$

- her expected benefit on $[0,\ \infty]$ is $E\left[Y_t dX_t\right] = rY_t a dt$

- the balance is $r\left(aY_t - h\left(a\right)\right)$

So, for the contract to be IC, we must have:

$$
Y_t A_t - h\left(A_t\right) = \max_{0 \leq a}\left\{aY_t - h\left(a\right)\right\} \quad \text{a.e} \qquad (7)
$$

The preceding argument is heuristic. Moreover, it gives only a necessary condition. It turns out that it can be made rigorous, and that in addition the condition is also sufficient:

Theorem 5 *The contract (A_t, C_t) is IC if and only if condition (7) holds*

Proof. Suppose (A, C) does not satisfy condition (7). Then there is an alternative contract (A^*, C^*) with:

$$
\begin{aligned}
Y_t A_t^* - h\left(A_t^*\right) &\geq Y_t A_t - h\left(A_t\right) \quad \text{a.e} \\
P\left[Y_t A_t^* - h\left(A_t^*\right)\right] &> P\left[Y_t A_t - h\left(A_t\right)\right]
\end{aligned}
$$

The agent picks $t > 0$ and plans to apply A^* for $s \leq t$ and A for $\geq t$. Her expected utility at t, conditional on $Z_{:t}$, is:

$$\frac{1}{r}V_t^* = E\left[\int_0^\infty e^{-rt}\left(u\left(C_t\right) - h_t\right)ds \mid \mathcal{F}_t^Z\right]$$

$$= \int_0^t e^{-rs}\left(u\left(C_s\right) - h\left(A_s^*\right)\right)ds + e^{-rt}W_t\left(A, C\right)$$

$$= W_0\left(A, C\right) + \int_0^t e^{-rs}\left(h\left(A_s\right) - h\left(A_s^*\right) - Y_s A_s + Y_s A_s^*\right)ds + \int_0^t e^{-rs}Y_s\left(dX_s - A_s^* ds\right) \quad (8)$$

Since $dX_s = A_s^* dt + \sigma dW_s$ for $s \leq t$, the last term is a martingale. Hence:

$$\mathbb{E}\left[V_t^*\right] = W_0\left(A, C\right) + \mathbb{E}\left[\int_0^t e^{-rs}\left(h\left(A_s\right) - h\left(A_s^*\right) - Y_s A_s + Y_s A_s^*\right)ds\right]$$

The integrand is non-negative, and positive on a set of positive measure in (t, ω). It follows that there is some $\bar{t}$ such that $\mathbb{E}\left[V_t^*\right] > W_0\left(A, C\right)$. But this means that switching from A^* to A at time $\bar{t}$ is better than sticking with A from the beginning. So (A, C) cannot be (IC). This proves that condition (7) is necessary.

Conversely, if (A, C) satisfies condition (7), then, by formula (8), V_t^* is a supermartingale, with:

$$W_0\left(A, C\right) = V_0^* \geq E\left[V_\infty^*\right] = W_0\left(A^*, C\right)$$

so A_t is at least as good as any alternative strategy A_t^*. This proves that condition (7) is sufficient. ∎

3.2 Reduction to an optimal control problem

Substituting condition (7) in (5) or (6), and assuming, for the sake of commodity, that the maximum is not attained at a boundary, so that $Y_t = h'\left(A_t\right)$, we get:

$$\frac{1}{r}dW_t = \left(W_t - u\left(C_t\right) + h\left(A_t\right)\right)dt + h'\left(A_t\right)\sigma dZ_t \quad (9)$$

$$= \left(W_t - u\left(C_t\right) + h\left(A_t\right) - h'\left(A_t\right)A_t\right)dt + h'\left(A_t\right)dX_t \quad (10)$$

Conversely, if (9) holds, we have $Y_t = h'\left(A_t\right)$, so that the contract (C_t, A_t) is IC. Sannikov's idea is to read this in another way, namely as a stochastic differential equation for a state variable W_t, controlled by C_t and A_t. The principal will then solve the optimal control problem:

$$\max_{C_t, A_t} r\mathbb{E}\left[\int_0^\infty e^{-rt}\left(dX_t - C_t dt\right)\right]$$

$$\frac{1}{r}dW_t = \left(W_t - u\left(C_t\right) + h\left(A_t\right)\right)dt + h'\left(A_t\right)\sigma dZ_t, \quad W_0 = w_0$$

(w is exogenous), the solution of which will be found in feedback form $C_t = c\left(W_t\right)$ and $A_t = a\left(W_t\right)$, for suitable functions $c\left(w\right)$ and $a\left(w\right)$. The contract will then be as follows:

8

1. The principal will pay the agent $C_t := c(W_t)$, where W_t is defined by (10) and A_t is defined below

2. The agent will perfom $A_t := a(W_t)$, where W_t is defined by (9)

This optimal contract is IC, by construction, so the agent will actually perform A_t, as stipulated in the contract, and formulas (10) and (9) coincide. So both parties will agree on the current value of W_t, which serves as a *performance index* for the agent. It depends on past performance, through formulas (10) and (9), and payments will be conditional on the current value of W_t. In other words, the contract will not be Markovian w.r.t. the revenue X_t, but it will be Markovian w.r.t. the performance index W_t.

Before we go any further, let us introduce a special class of contracts. Suppose $C_t = c$ for time $t \geq T$: the principal pays a constant compensation after time T. As we noticed earlier, the agent then stops exerting effort, so that $A_t = 0$ for $t \geq T$. In fact, the principal pensions off the agent. It is not immediately obvious why he should wish to pay her for doing nothing, as he has always the possibility to fire her, i.e. to set $C_t = 0$. We will see, however, that the optimal contract involves retiring the agent instead of firing her.

The time T when the agent is pensioned off must be specified in the contract, so it is a stopping time w.r.t. $\mathcal{F}_t^X$. The continuation value at time T is easily computed:

$$W_T = r \int_T^\infty e^{-rt} u(c)\, ds = u(c)$$

so $c = u^{-1}(W_T)$. The value of the contract for the principal is given by:

$$r\mathbb{E}\left[\int_0^T e^{-rt}(dX_t - C_t dt) + e^{-rT}\int_T^\infty e^{-rt} cds\right] =$$

$$r\mathbb{E}\left[\int_0^T \left(e^{-rt}(A_t - C_t)\, dt + \sigma dZ_t\right) + \frac{e^{-rT}}{r} u^{-1}(W_T)\right]$$

We now summarize. The principal has to solve the stochastic optimal control problem:

$$\sup\left\{r\mathbb{E}\left[\int_0^T \left(e^{-rt}(A_t - C_t)\, dt + \sigma dZ_t\right) + \frac{e^{-rT}}{r} u^{-1}(W_T)\right]\right\}$$

$$\frac{1}{r} dW_t = (W_t - u(C_t) + h(A_t))\, dt + h'(A_t)\, \sigma dZ_t$$

The optimisation is carried over the controls $A_t \in [0, \bar{a}]$ and $C_t \geq 0$, and the stopping time $T \geq 0$. The state variable is W_t. The initial value, $W_0 = w_0$, is exogenous: it is part of the contract, and the principal will choose it in the way which is best for himself.

3.3 The HJB equation.

As usual, to solve the optimal control problem, we introduce the value function

$$F\left(w\right) = r\sup E\left[\int_0^T e^{-rt}\left(dX_t - C_t dt\right) - \frac{e^{-rT}}{r}u^{-1}\left(W_T\right) \mid W_0 = w_0\right]$$

$F\left(w\right)$ is the highest expected utility the principal can obtain while guaranteeing the agent an expected utility of w.

$F : [0, \quad \infty) \to \mathbb{R}$ is continuous, and $F\left(w\right) \geq -u^{-1}\left(w\right)$ everywhere. T is the first time when $F\left(W_t\right) \leq -u^{-1}\left(w\right)$ The function F satisfies the Hamilton-Jacobi-Bellman equation, which in this case is a *quasi-variational inequality*:

$$F\left(w\right) = \max_{\substack{\bar{a} \geq a \geq 0, \\ c \geq 0}}\left\{\begin{array}{c} -u^{-1}\left(w\right), \\ a - c + F'\left(w\right)\left(w - u\left(c\right) + h\left(a\right)\right) + \frac{r}{2}F''\left(w\right)h'\left(a\right)^2\sigma^2 \end{array}\right\}$$

$$(11)$$

Denote by $A\left(w\right)$ and $C\left(w\right)$ the points a and c where the maximum is attained on the RHS.

Theorem 6 *Suppose F solves (HJB) with $F\left(0\right) = 0$. Pick some w_0 and define W_t as follows:*

$$\frac{1}{r}dW_t = W_t - u\left(C\left(W_t\right) + h\left(A\left(W_t\right)\right)\right)dt \ + h'\left(A\left(W_t\right)\right)dZ_t \qquad (12)$$

with $W_0 = w_0$ Then the contract $C_t = C\left(W_t\right)$, $A_t = A\left(W_t\right)$ is IC, IR, and has value w_0 for the agent and $F\left(w_0\right)$ for the principal. The principal pensions off the agent at time $T := \inf\left\{t \mid F\left(W_t\right) \leq -u^{-1}\left(W_t\right)\right\}$. Any IC and IR contract starting from $W_0 = w$ yields to the principal a profit less than or equal to $F\left(w\right)$

Note that the stopping time T occurs either when $W_t = 0$ or when $W_t = \bar{w}$, where $\bar{w}$ is the smallest positive solution of $F\left(w\right) = -u^{-1}\left(w\right)$. It turns out that it is finite: $T < \infty$ a.s.

Proof. Let us prove that the contract yields value $F\left(w_0\right)$ for the principal. By construction, W_t is the continuation value for the agent, so $W_0 = w_0$ is the expected utility the agent derives from the contract when she enters it. The expected utility to the principal is:

$$r\mathbb{E}\left[\int_0^T e^{-rs}\left(A_s - C_s ds\right) - \frac{e^{-rT}}{r}u^{-1}\left(W_T\right)\right]$$

The random variable $G_{t:} := \int_0^t e^{-rs}\left(A_s - C_s ds\right) + r^{-1}e^{-rt}F\left(W_t\right)$ is a diffusion. Applying Ito's Lemma to 12 and using (HJB), we check that it is a martingale. By the optional stopping theorem,

$$E\left[G_T\right] = G_0 = F\left(W_0\right)$$

One then checks that this contract is optimal, provided the solution F is convex. We refer to [5] for a full proof. ∎
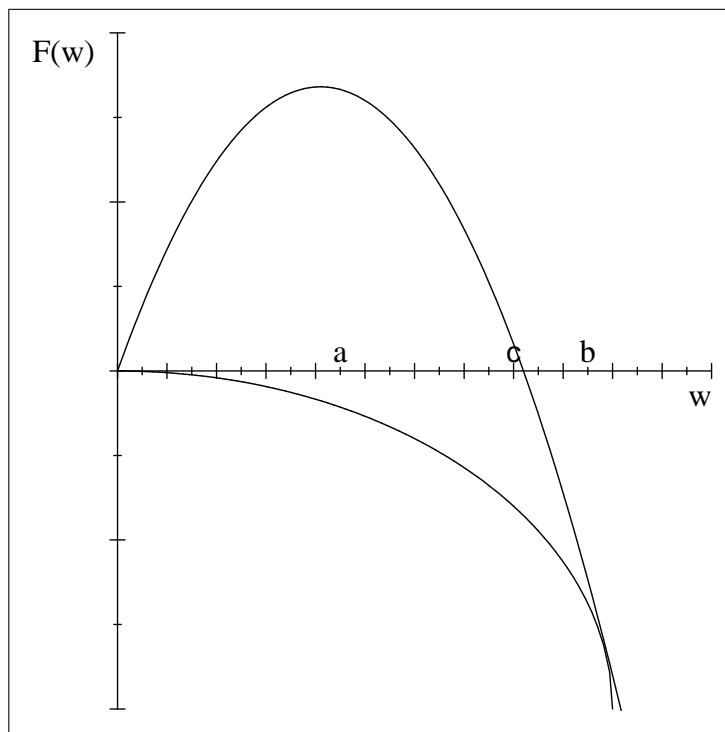
Figure 1: the solution of HJB

The solution is depicted in Figure 1. The upper curve is $F(w)$, the lower curve is $F_0(w) := -u^{-1}(w)$. The two curves meet at $w = 0$ and $w = b$, and by the smooth pasting principle (which can be derived rigorously from (HJB) itself), the two curves much have a common tangent at b. The principal will choose $w_0 = a$, where $F(a) = \max_w F(w)$, and offer the corresponding contract to the agent. The horizontal coordinate $w = W_t$ then evolves according to equation (12), starting from $W_t = a$, until it meets either the left boundary 0 or the right b, which will occur a.s. at some finite time T. If $W_T = 0$, the principal fires the agent ($C_t = 0$ and $A_t = 0$ for $t \geq T$), if $W_T = b$, the principal pensions the agent off ($C_t = u^{-1}(b)$ and $A_t = 0$ for $t \geq T$). Further analysis shows that $A(w)$ decreases and $C(w)$ increases when $w \to b$ from the left: as the agent nears the retirement point, she works less and less and gets paid more and more.

Let us denote by c the point where $F(w) = 0$. In the region where $c < w \leq b$, we have $F(w) < 0$, meaning that the principal is expecting to lose money. To test the reader's understanding of the situation, here are three questions:

1. Why does the principal continue the contract in the region $c < w \leq b$, and not quit as soon as c is reached ?

2. Why does the principal pension off the agent when b is reached, instead of simply firing her ?

3. Why does the principal fire the agent when 0 is reached, instead of offering her to start again at a, which would make both of them better off ?

The answers is all about incentives. The principal is bound by the contract, so if he wants to do any one of 1, 2, or 3, he must write it in the contract, so the agent will know it as soon as she signs in. Then:

1. If the agent knows that the principal is going to quit at c, she will try not to reach c, because she is making positive profit in the meantime, so she will slack off in the region $[0, c]$, thereby hurting the profits of the principal

2. If the agent knows she will be fired when b is reached, she will try not to reach b, because she is making positive profits in the meantime, so she will slack off in the region $[c, b]$, where the principal is losing money, and keep him there for ever.

3. If the agent knows she will not be fired when 0 is reached, but simply offered a new contract, she will not be afraid of reaching 0, and slack in the region $[0, a]$, thereby hurting the profit of the principal.

The last situation is particulary interesting: it is called renegociation. It is like promising your child he will be punished if he breaks grandmother's vase, which has been defacing the living-room for twenty years. When the vase is broken, you are rather happy that it is gone, you do not want to punish the child, and he is no glutton for punishment either. So, even though both of you would be better off not doing it, still you punish the child, because if he knows ahead of time that he will get away with it, your word will carry no authority.

4 Conclusion

Moral hazard and limited liability are ubiquitous in our society. We have investigated their consequences in a very simple situation. The contract we get at the end is quite complicated, and one even counter-intuitive: the principal has to pay the agent more and more to do less and less, and when he can no longer afford her (diminishing) services, he has to pension her off comfortably ! However, no one can deny that it bears a close resemblance to what happens in the finance industry nowadays: traders routinely get huge bonuses, bankers who have led their institution to bankruptcy and lost incredible amounts of investors' money are pensioned off with golden parachutes. Never before in the history of the world has incompetence been so highly rewarded.

Moral hazard and limited liability occur in many other situations as well. Here are just some of them:

- How can the owners of a firm induce the managers to spend enough effort in preventing accidents, in an industry where accidents are rate but very costly (to the owners) ? See [1]

- What kind of contracts should financiers pass with entrepreneurs, when the latter have the possibility not to exert due diligence, or even to divert investment money to other purposes ? See [2]

- All these (second-best) contracts turn out not to be renegociation-proof. Can one devise incentive-compatible contracts which are renogociation-proof (third-best) ? See [4]

In all these models, the message is the same: if you cannot punish the agent (limited liability), and cannot monitor what she is doing (moral hazard), all you can do is pay her so much that she will feel she owns the project. This is what banks and firms do nowadays, by giving their executives stock options and bonuses, not to mention ever-increasing paychecks.

I would like to conclude with one word of caution to the reader: this is not a law of nature, or even a law of economics. This is the logical consequence of the rules we have set, namely moral hazard and limited liability. If you do not like the consequences, if you do not like traders getting all this money for risking other peoples' savings, or bankers getting huge bonuses or golden parachutes when their institution has been rescued by taxpayers, then you have to change the rules of the game. Moral hazard can be reduced by monitoring, and limited liability does not hold when negligence can be proven. Unfortunately, this would require society to invest much more in monitoring and regulation agencies than it does at present; from present trends, I do not think we are going that way. Regulatory bodies continue to be the "parent pauvre" of the financial industry, while the best and the brightest flock to business schools, where they are taught precisely what I have described in this paper, and apply their knowledge to the benefit of their own career.

References

[1] Biais, Mariotti, Rochet, Villeneuve, "*Large risks, limited liability and dynamic moral hazard*", EMA (2010), 73-118

[2] See Biais, Mariotti, Plantin, Rochet, "*Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications*", The Review of Economic Studies (2007) p. 345-390.

[3] Cvitanic and Zhang, "*Contract theory in continuous-time models*", Springer, 2012

[4] Ekeland and Rochet, work in progress

[5] Sannikov, "*A continuous-time version of the principal-agent problem*", RES (2008) 75, 957-984

[6] Sannikov, "*Contracts: the theory of dynamic principal-agent relationships and the continuous-time approach*", Working paper, 2012