Motivating Innovation with a Structured Incentives Scheme under Continuous States

Chengli Zheng[1*] and Yan Chen[2]
School of Economics, Huazhong Normal University, China[1]
Department of Applied Mathematics & Statistics, Stony Brook University, USA.[2]
zhengchengli168@163.com[1], stchenyan88@163.com[2]
**Corresponding Author*

Received 16 August 2014; received in revised form 27 October 2014; accepted 10 November 2014

Abstract

The problem of incentive is an important component of the separation of ownership and control. A large amount of literature focuses on the problem of how to use pay-for-performance schemes to both inspire agents to exert effort and to deter agent-based resource tunneling. Manso (2011) proposes the use of structured incentive schemes with two periods to motivate innovation under discrete states. In combining these two perspectives, this paper proposes a version with continuous states and points out that an agent can, simultaneously, innovate while exerting effort to obtain greater output per unit time. By being offered a suitable incentive contract, the agent will carry out the exploration action plan, although he may fail. In the meantime, he will exert all his efforts to raise production, which determines his reward.

Keywords: Motivating innovation, structured incentive scheme, exploitation and exploration, continuous state

1. Introduction

Since Berle and Means (1932) pointed out drawbacks with the separation of ownership and control, the incentive issue has become a subject of interest for this field. Harris and Raviv (1978) and Holmstrom (1979) have mostly focused on the problems of how to inspire an agent to exert effort or deter agents from tunneling resources away from the corporation by applying principal-agent models. Manso (2011) presents a different view. He studied how to build a certain incentive structure to motivate the agent to be more innovative with a two-period model. He showed that incentive schemes that motivate innovation should be structured differently from standard pay-for-performance schemes used to induce effort or avoid tunneling. Innovation involves the exploration of new untested approaches that are likely to fail. Therefore, standard pay-for-performance schemes that punish failures with low rewards and termination may in fact have an adverse effect on innovation. In contrast, an optimal incentive scheme that motivates innovation exhibits substantial tolerance (or even reward) for early failure and reward for long-term success. Under this incentive scheme, compensation depends not only on the total performance overall, but also on the path of the performance; an agent who performs well initially but poorly later earns less than an agent who performs poorly initially but well later or even an agent who performs poorly repeatedly.

Based on the framework of Manso (2011), this paper studies the incentives for innovation with non-fixed rewards for the agent. Our model absorbs the advantages of the two aforementioned directions: incentive schemes for motivating innovation and standard pay-for-performance schemes. We give the standard for success, and the reward of the agent depends on the amount of the excess output over the baseline. The fixed wage and non-fixed wage (wage rate) are designed. These structured incentives can motivate the agent to select a more innovative work method and stimulate the agent to exert effort to get a better output in the meantime. The reward for the agent is comprised of two parts: one fixed part which is independent of any situation, and another non-fixed part which depends on the output. The fixed part can mainly be used to tolerate the failure of the exploration, and the non-fixed part is used to stimulate the agent to engage in innovative action and to exert all his efforts to get the best reward.

Similar to Manso (2011), we use a two-period innovation process to deal with the incentive problem. To model the innovation process, we use a class of Bayesian decision models known as bandit problems. We focus on the central concern that arises with bandit problems: the tension between the exploration of new untested actions and the exploitation of well-known actions. For the related literature see Holmstrom (1989), Aghion and Tirole (1994), Arrow (1969), March (1991), Moscarini and Smith (2001), Hellmann and Thiele (2011), Tian and Wang (2014), Ederer and Manso (2013) and other literature cited in Manso (2011). However, there are differences here, too. The model of Manso (2011) just considers two states: success and failure, and the optimal contracts depend only on the probability of success or failure, not on the amount of the outputs. Our model is treated under the continuous states, and the optimal contracts depend on the distribution of the production, not only on the probability of success or failure, but also on the amount of the outputs.

The rest of the paper is arranged as follows: section Π examines the bandit problem for tension between exploration and exploitation; section Ш presents the principal-agent problem regarding tension; section IV gives the solutions to the principal-agent problem, namely optimal incentive contracts for exploration and exploitation, respectively; and the last section concludes the paper.

2. Examining the Bandit Problem for Tension between Exploration and Exploitation

Here, we review the two-armed bandit problem with the one known arm as per Manso (2011) and Zheng and Chen (2013). This illustrates the tension between exploration and exploitation. Exploitation is a well known action, and the agent can receive a reasonable payoff clearly with little cost. However, exploration is a new untested action with a high uncertainty of success. If the agent takes exploratory action and has a success, the output can be very high, but it is more likely to fail and cost more. Basically, the principal expects the agent to take exploratory action, but the agent wants to take the exploitation action. Consequently, there is tension! To solve this problem, a structured incentive scheme must be designed. The original models were proposed under discrete states. We extend these to be one model with continuous states.

We assume that the agent lives for only two periods (More periods can be assumed but the results may be different. Here, we only consider two periods. One reason is this will show some basic insights to this problem; another reason is that the model is not too complex to be treated. The long term will be checked in our future work). In each period, $t \epsilon T = \{1,2\}$, the agent takes an action $i \in I$, producing output R_{ti}, which is a random variable with a cumulative distribution function $F_{R_{ti}}(x) = P[R_{ti} \le x]$. The principal gives the baseline B_t of the output for each period $t \in T$ to evaluate the performance of the agent. If $R_{ti} > B_t$, the agent is judged as a "success"; if $R_{ti} \le B_t$ the agent is judged as a "failure". The cumulative distribution function $F_{R_{ti}}(x)$ may be unknown for some of the actions. To obtain information about $F_{R_{ti}}(x)$ for these actions, the agent needs to engage in experiments during the first period. We let $h(R_{ti})$ denote the return function on output R_{ti}. We also let $E[h(R_{ti})]$ denote the unconditional expectation of $h(R_{ti})$, let $E[h(R_{ti})|R_{t-1j} > B_{t-1}]$ denote the conditional expectation of $h(R_{ti})$ given a success on action j in the last period, and $E[h(R_{ti})|R_{t-1j} \leq B_{t-1}]$ denote the conditional expectation of $h(R_{ti})$ given a failure for action j in the last period. When the agent takes action $i \in I$ in period $t \in T$, he only learns about the information for the distribution of R_{t+1i} for the next period, so that

$$E[h(R_{ti})] = E[h(R_{ti})|R_{t-1j}] \text{ for } i \neq j$$

This means that if the agent wants to know the information for the distribution of R_{t+1i} for the next period, he must engage in the experiment of action i with unknown distribution in this period.

Because there is no new information for the unconditional expectation of $h(R_{ti})$, namely, it is independent of time, so we denote $E[h(R_{ti})] = E[h(R_i)]$ in this situation.

Our main focus of interest is on the tension between two actions: action 1 is exploration and action 2 is exploitation. We assume that in each period $\in T$, the agent chooses between these two actions. Action 1 is the conventional work method, has a known distribution of R_{t1} in any period $t \in T$, namely $R_{t1} = R_1$, such that $E[h(R_{t1})] = E[h(R_{t1})|R_{t-11}] = E[h(R_1)]$

Action 2 is the new work method, has an unknown distribution of R_{t2} such that[1]
$E[h(R_{t2})|R_{t-12} \leq B_{t-1}] < E[h(R_{t2})] < E[h(R_{t2})|R_{t-12} > B_{t-1}]$

This means that if the agent has a success with the new work method, then he updates his belief that there is further possibility that the new work method will succeed. Or, if the agent observes a failure with the new work method, then he updates his beliefs that there is more possibility that the new work method will fail.

[1] Here we assume that $h(R_{t2})$ is increasing the function on R_{t2}.

We assume that Action 2 has an exploratory nature. This means that when the agent experiments with the new work method, he is, initially, not as likely to succeed as when he conforms to the conventional work method. However, if the agent observes a success with the new work method, then he updates his beliefs about the probability of success with the new work method, so that the new work method is perceived as being better than the conventional work method. This is captured as follows:
$E[h(R_2)]<E[h(R_1)]<E[h(R_{t2})|R_{t-12}>B_{t-1}]$

In fact, the agent may shirk and not choose either of the two work methods mentioned above. This action 0 is allowed in the model. Shirking has zero private cost, but has a lower expected return than either of the two work methods. Here, we assume that action 0 (shirking) has a return R_0 with a known distribution in any period $t \in T$. Without losing generality, we assume that there exists a stochastic dominant relationship as follows:

$$\left(R_{t2}|R_{t-12}>B_{t-1}\right) \overset{FSD}{\succ} R_1 \overset{FSD}{\succ} R_2 \overset{FSD}{\succ} \left(R_{t2}|R_{t-12}\leq B_{t-1}\right) \overset{FSD}{\succ} R_0$$

Where $X \overset{FSD}{\succ} Y$, it means that X stochastically dominates Y in the first order, namely $F_X(\eta) \leq F_Y(\eta)$, for all $\eta \in R$.

So, if $h(\bullet)$ is a non-decreasing function, we have

$$E[h(R_0)] < E[h(R_{t2})|R_{t-12} \leq B_{t-1}] < E[h(R_2)] < E[h(R_1)] < E[h(R_{t2})|R_{t-12} > B_{t-1}] \quad (1)$$

In fact, the model is a three-armed bandit problem, namely = {0,1,2}, but we only consider the tension between exploration and exploitation. The agent is risk-neutral and has a discount factor normalized to one. The agent thus chooses an action plan $\langle i_k^j \rangle$ to maximize his total expected payoff. Where $i \in I$ is the first-period action, $j \in I$ is the second-period action if there is success in the first period; $k \in I$ is the second-period action if there is failure in the first period.

Two action plans need to be considered. Action plan $\langle 1_1^1 \rangle$, which Manso (2011) called exploitation, is just the repetition of the conventional work method. Action plan $\langle 2_1^2 \rangle$, which Manso called exploration, is to initially try the new work method, sticking to the new work method if there is success in the first period, and revert to the conventional work method if there is failure in the first period. Apparently, the total payoff for action plan $\langle 2_1^2 \rangle$ from exploration is higher than that of action plan $\langle 1_1^1 \rangle$ from exploitation if, and only if, $E[R_2] > E[R_1] - E\{1_{R_{12}>B_1}(E[R_{22}|R_{12} > B_1] - E[R_1])\}$

When the agent tries the new work method, he obtains information about R_{t2}. This information is a useful guide for the agent's decision in the second period, since the agent can switch to the conventional work method if he ascertains that the new work method is not worth pursuing. The agent may thus be willing to try the new work method even though the initial expected return $E[h(R_2)]$ with the new work method is lower than the expected return $E[h(R_1)]$ with the conventional work method.

3. The Principal-Agent Problem

In this section, we introduce incentive problems to the three-armed bandit problem with the two known arms as reviewed in the previous section.

The principal hires an agent to perform a task described in the previous section. In each period, the agent incurs private costs $C_i \geq 0$ if he takes action $i = 1,2$, but can avoid these private costs by taking action $i = 0$, shirking ($c_0 = 0$).

We assume that the principal does not observe the actions taken by the agent. As such, before the agent starts working, the principal offers the agent a contract $\langle \bar{\lambda}, \bar{w} \rangle = \{\langle \lambda_1, w_1 \rangle, \langle \lambda_2, w_2 \rangle, \langle \lambda_3, w_3 \rangle\}$ that specifies the agent's wages contingent on future performance. The agent has limited liability, meaning that his wages cannot be negative. Here, w_s $(s = 1,2,3)$ is a fixed wage, which is the minimum wage in any situation, and λ_s is the wage rate for extra returns in the situation of success. This means that if the action is a failure, the agent will still get a fixed wage w_s; if it is a success, he will get a fixed wage w_s plus a flexible wage $\lambda_s(R_s - B_s)1_{R_s - B_s}$. Specifically, $\langle \lambda_1, w_1 \rangle$ is the wage rate and the fixed wage in the first period, respectively. $\langle \lambda_2, w_2 \rangle$ is the wage rate and the fixed wage in the second period on condition of success in the first period, respectively. $\langle \lambda_3, w_3 \rangle$ is the wage rate and fixed wage in the second period conditional on failure in the first period, respectively.

Different from that of Manso (2011), the contract $\langle \bar{\lambda}, \bar{w} \rangle$ in our model is not a fixed wage. While a fixed wage in the case of failure, $\bar{\lambda}$ is a fixed wage rate in the case of success. When the agent succeeds in one period t, according to the baseline of success B_t given by the principal in advance, he will get a payoff w_s plus $\lambda_s(R_s - B_s)1_{R_s - B_s}$, $s = 1,2,3$, which is dependent on the output. The more output it produces, the more wage rewards he gets. So, the contract $\langle \bar{\lambda}, \bar{w} \rangle$ for our Principal-agent model has two functions: one is to motivate the agent to be more innovative and the other is to inspire the agent to exert more effort.

This is different from that of Zheng and Chen (2013), where the w_s is not a minimum wage, which may lead to the situation where the wage for success will be lower than that for failure. Here, we revise this fault.

In addition to these differences, another feature is that the models here are built with continuous states. To illustrate the process of the reward structure, see figure 1 as follows.

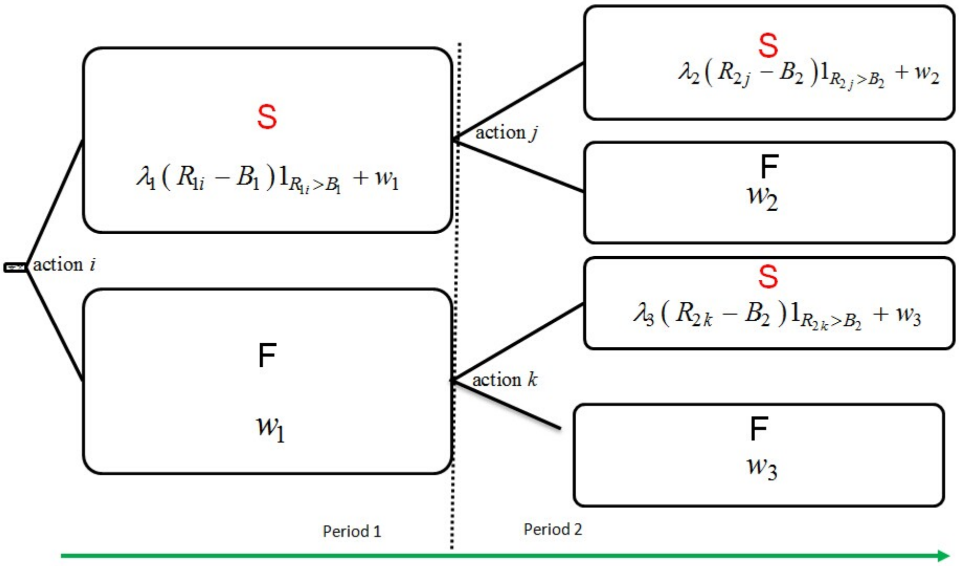

Figure 1: Structured Reward Action Plan $\langle i_k^j \rangle$
S-success, F-failure

We assume that both the principal and the agent are risk-neutral and have a discount factor of one, just for simplicity. When the principal offers the agent a contract $\langle \bar{\lambda}, \bar{w} \rangle$ and the agent takes on the action plan $\langle i_k^j \rangle$, the total expected payments from the principal to the agent are given by

$$W(\vec{\lambda}, \vec{w}, \langle i_k^j \rangle) = E[\lambda_1(R_{1i} - B_1)1_{R_{1i}>B_1} + w_1] + E\left\{1_{R_{1i}>B_1} E\left[\lambda_2(R_{2j} - B_2)1_{R_{2j}>B_2} + w_2 | R_{1i} > B_1\right]\right\} + E\{1_{R_{1i} \le B_1} E[\lambda_3(R_{2k} - B_2)1_{R_{2k}>B_2} + w_3 | R_{1i} \le B_1]\} \quad (2)$$

Apparently, the model of Manso (2011) and Zheng and Chen (2013) are special discrete cases of our model.

Because $E[(R_{ti} - B_t)1_{R_{ti}>B_t}]$ can be viewed as a call option whose underlying asset is output R_{ti} and strike price is B_t, we denote $op_{ti} = E[(R_{ti} - B_t)1_{R_{ti}>B_t}]$.

Similarly, we denote $\overline{op}_{2j}^{1i} = E\left[(R_{2j} - B_2)1_{R_{2j}>B_2} | R_{1i} > B_1\right]$, and $\underline{op}_{2k}^{1i} = E[(R_{2k} - B_2)1_{R_{2k}>B_2} | R_{1i} \le B_1]$.

So equation (2) can be rewritten as

$$W(\vec{\lambda}, \vec{w}, \langle t_k^j \rangle) = \lambda_1 op_{1i} + w_1 + E\left\{1_{R_{1i}>B_1}\left(\lambda_2 \overline{op}_{2j}^{1i} + w_2\right)\right\} + E\left\{1_{R_{1i} \le B_1}\left(\lambda_3 \underline{op}_{2k}^{1i} + w_3\right)\right\} \quad (3)$$

It means that the total expected payments are comprised of a series of options.

According to the assumptions in the previous section, we have

$$op_{t0} < \underline{op}_{22}^{12} < op_{t2} < op_{t1} < \overline{op}_{22}^{12} \quad (4)$$

When the agent takes on the action plan $\langle t_k^j \rangle$, the total expected costs incurred by the agent are given by

$$C(\langle i_k^j \rangle) = c_i + E[1_{R_{1i}>B_1}]c_j + E[1_{R_{1i} \le B_1}]c_k \quad (5)$$

Here, we consider a non-cooperative game (Stackelberg game). It needs to be pointed out that the model assumes a common knowledge framework in which all information is known to both agents. This assumption is due to the nature of the Stackelberg game. However, the problem here is a little different from the standard solution. We only want to know what kind of wage structure can encourage the agent to take on the objective action plan $\langle i_k^j \rangle$, such as the innovative action plan $\langle 2_1^2 \rangle$ or the conventional action plan $\langle 1_1^1 \rangle$.

We say that contract $\langle \bar{\lambda}, \bar{w} \rangle$ is an optimal contract that implements an action plan $\langle i_k^j \rangle$ if it minimizes the total expected payments from the principal to the agent,

$$W(\vec{\lambda}, \vec{w}, \langle i_k^j \rangle) \quad (6)$$

Subjected to the incentive compatibility constraints, $W(\vec{\lambda}, \vec{w}, \langle i_k^j \rangle) - C(\langle i_k^j \rangle) \geq W(\vec{\lambda}, \vec{w}, \langle l_n^m \rangle) - C(\langle l_n^m \rangle)$ $(IC_{\langle l_n^m \rangle})$.

This is a linear program with six unknowns and 27 constraints because $l, m, n \in I$. When more than one contract solves this program, we restrict attention to the contract that pays the agent earlier, as per Manso (2011).

The principal's expected profit from implementing action plan $\langle i_k^j \rangle$ is given by

$$\Pi(\langle i_k^j \rangle) = Y(\langle i_k^j \rangle) - W(\vec{\lambda}(\langle i_k^j \rangle), \vec{w}(\langle i_k^j \rangle), \langle i_k^j \rangle) \quad (7)$$

Where

$$Y(\langle i_k^j \rangle) = E[R_{1i}] + E\{1_{R_{1i} > B_1} E[R_{2j} | R_{1i} > B_1]\} + E\{1_{R_{1i} < B_1} E[R_{2k} | R_{1i} \leq B_1]\} \quad (8)$$

is the principal's total expected revenue when the agent uses action plan $\langle i_k^j \rangle$ and $\langle \vec{\lambda}(\langle i_k^j \rangle), \vec{w}(\langle i_k^j \rangle) \rangle$ is the optimal contract that implements action plan $\langle i_k^j \rangle$, the principal thus chooses action plan $\langle i_k^j \rangle$ that maximizes $\Pi(\langle i_k^j \rangle)$.

The assumptions for the principal-agent problem studied here are standard except that there is learning about the technology being employed. This gives rise to tension between the exploration and the exploitation, since there is nothing to be learned about the conventional technology, but a lot to be learned about the new technology.

4. Incentives for Exploration and Exploitation

Here we present the optimal contracts that implement exploration and exploitation.

4.1 Incentives for Exploitation

Recall from Section II that exploitation represented by action plan $\langle 1_1^1 \rangle$.

$$W(\vec{\lambda}, \vec{w}, \langle 1_1^1 \rangle) = \lambda_1 op_{11} + w_1 + E\left\{1_{R_{11} > B_1}\left(\lambda_2 \overline{op_{21}^{11}} + w_2\right)\right\} + E\left\{1_{R_{11} \leq B_1}\left(\lambda_3 \underline{op_{21}^{11}} + w_3\right)\right\} \quad (9)$$

Given the goal of the action plan $\langle 1_1^1 \rangle$, the principal must offer optimal contracts so that the agent implements the exploitation. The optimal contracts $\langle \bar{\lambda}, \bar{w} \rangle$ must maximize $\Pi(\langle 1_1^1 \rangle)$, namely minimize $W(\vec{\lambda}, \vec{w}, \langle 1_1^1 \rangle)$ subject to the incentive compatibility constraints, $W(\vec{\lambda}, \vec{w}, \langle 1_1^1 \rangle) - C(\langle 1_1^1 \rangle) \geq W(\vec{\lambda}, \vec{w}, \langle l_n^m \rangle) - C(\langle l_n^m \rangle)$ $IC_{\langle l_n^m \rangle}$.

We then derive the optimal contract that implements exploitation. The following definitions will be useful when stating Proposition 1:

$$\beta_0 = \frac{1}{1 + E[1_{R_2 > B_1}]}\left(\frac{E\left[1_{R_2 > B_1}\left(\overline{op_{22}^{12}} - op_{20}\right)\right]}{op_{21} - op_{20}} + \frac{op_{12} - op_{10}}{op_{11} - op_{10}}\right)$$

Because the distribution of return R_2 in the first period is unknown, we use the expectation of $E[1_{R_2 > B_1}]$ to denote it. We also denote $p_0 = E[1_{R_0 > B_1}]$, $p_1 = E[1_{R_1 > B_1}]$ directly.

PROPOSITION 1: The optimal contract $\langle \vec{\lambda}, \vec{w} \rangle_1^*$ that implements exploitation is such that $w_1 = w_2 = w_3 = 0$, $\lambda_2 = \lambda_3 = \frac{c_1}{op_{21} - op_{20}}$ $\lambda_1 = \frac{c_1}{op_{11} - op_{10}} + \frac{(1 + E[1_{R_2 > B_1}])c_1}{op_{11} - op_{12}}\left(\beta_0 - \frac{c_2}{c_1}\right)^+$ where $(x)^+ = max(x, 0)$.

The formal proofs for each of the propositions are omitted and limited to the length. However, the main intuition behind Proposition 1 is as follows. To implement exploitation, the principal must prevent the agent from both shirking and exploring. If c_2 is high when relative to c_1, only shirking constraints are binding. Therefore, the optimal contract that implements exploitation is similar to the optimal contract used to induce the agent to exert effort in a standard work-shirk principal-agent model. If c_2 is low when relative to c_1, the exploration constraint is binding. To prevent exploration, the principal must pay the agent an extra premium if there is success in the first period. This extra premium decreases in c_2/c_1, because when c_2/c_1 increases, the agent becomes less inclined to explore.

Similarly, the baseline B_t will affect the result. If $B_1 \geq B_2$, then $\lambda_1 \geq \lambda_2 = \lambda_3$. This can be interpreted as when the baseline standard for success decreases, the difficulty for success in the second period decreases, and the exploration constraint may be binding. To prevent exploration, the principal must pay the agent an extra premium if there is success in the first period. However, if $B_1 < B_2$, the difficulty for success in second period increases, and the exploitation constraint may be binding, indicating that the principal may not need to pay the agent an extra premium if there is success in the first period. It means that the following $\lambda_1 < \lambda_2 = \lambda_3$ may hold at this time.

To encourage the agent to use the conventional method, there are no fixed minimum wages. This means that failure is not tolerated during the whole process.

Proposition 1 is for the optimal incentive contract for the exploitation plan $\langle 1_1^1 \rangle$, which is just a comparison and complements the optimal incentive contract for the exploration plan $\langle 2_1^2 \rangle$ (proposition 2 in the next subsection). From proposition 1, the differences between exploitation and exploration can be checked. For the principal, the action plan he wants the agent to undertake are either $\langle 2_1^2 \rangle$ or $\langle 1_1^1 \rangle$. Therefore, it is important that the incentive schemes for these two plans (proposition 1 and 2) are displayed here.

4.2 Incentives for Exploration

Proposition 2 derives the optimal contract that implements exploration. Recall from Section II that exploration is given by action plan $\langle 2_1^2 \rangle$.

$$W(\vec{\lambda},\vec{w},\langle 2_1^2 \rangle) = \lambda_2 op_{12} + w_1 + E\left\{1_{R_{12}>B_1}\left(\lambda_2 \overline{op_{22}^{12}} + w_2\right)\right\} + E\left\{1_{R_{12}\leq B_1}\left(\lambda_3 \underline{op_{21}^{12}} + w_3\right)\right\} \quad (10)$$

Given the goal of action plan $\langle 2_1^2 \rangle$, the principal must offer the optimal contracts that implement the exploration. The optimal contracts $\langle \vec{\lambda},\vec{w} \rangle$ must maximize $\prod(\langle 2_1^2 \rangle)$, namely minimize $W(\vec{\lambda},\vec{w},\langle 2_1^2 \rangle)$ subject to the incentive compatibility constraints, $W(\vec{\lambda},\vec{w},\langle 2_1^2 \rangle) - C(\langle 2_1^2 \rangle) \geq W(\vec{\lambda},\vec{w},\langle l_n^m \rangle) - C(\langle l_n^m \rangle)$ $(IC_{\langle l_n^m \rangle})$.

The form of the optimal contract that implements exploration will depend on whether exploration is moderate or radical. In the following definition, we classify the exploration into two types: moderate and radical. The reason for this classification is that when we solve the optimization, there are significant differences in the optimal contract, especially for w_1, which is dependent on the possibility of failure and rewards success. This can be described by the following definition.

DEFINITION 1: Exploration is radical if $\frac{E[1_{R_2\leq B_1}]}{E[1_{R_1\leq B_1}]} \geq \frac{E[1_{R_2>B_1}\overline{op_{22}^{12}}]}{E[1_{R_1>B_1}]op_{21}}$ but moderate otherwise.

Exploration is radical if the likely ratio between exploration and exploitation of a failure in the first period is greater than the reward ratio between the exploration and exploitation of two consecutive successes. We call this exploration radical one because it has a high expected probability of failure in the first period relative to the probability of failure regarding the conventional action. Alternatively, we call it moderate exploration because it has a lower expected probability of failure in the first period relative to the probability of failure of the conventional action. Apparently, the incentives for the two types of exploration are different.

The following definitions will also be useful when stating Proposition 2:

$$\beta_1 = \frac{E\left[1_{R_2>B_1}\left(\overline{op_{22}^{12}} - op_{20}\right)\right]}{(1 + E[1_{R_2>B_1}])(op_{21} - op_{20})}$$

$$\beta_2 = \beta_1 + \frac{1}{1 + E[1_{R_2>B_1}]} \frac{E\left[1_{R_2>B_1}\overline{op_{22}^{12}} - p_0 op_{20}\right]}{(p_1 - p_0)op_{21}}$$

PROPOSITION 2: The optimal contract $\langle \vec{\lambda},\vec{w} \rangle_2^*$ that implements exploration is such that $\lambda_1 = 0$, $\lambda_3 = \frac{c_1}{op_{21} - op_{20}}$ and $w_2 = w_3 = 0$

If exploration is moderate, then $w_1 = 0$ and

$$\lambda_2 = \frac{c_1}{op_{21}-op_{20}} - \frac{(1+E[1_{R_2>B_1}])c_1}{E\left[1_{R_2>B_1}\overline{op_{22}^{12}}\right]-p_0 op_{20}}\left(\beta_1 - \frac{c_2}{c_1}\right)^+ + \frac{(1+E[1_{R_2>B_1}])c_1}{E\left[1_{R_2>B_1}\overline{op_{22}^{12}}\right]-p_0 op_{21}}\left(\frac{c_2}{c_1} - \beta_1\right)^+ + \frac{(1+E[1_{R_2>B_1}])(p_1-p_0)op_{21}c_1}{\left(E\left[1_{R_2>B_1}\overline{op_{22}^{12}}\right]-p_1 op_{21}\right)\left(E\left[1_{R_2>B_1}\overline{op_{22}^{12}}\right]-p_0 op_{21}\right)}\left(\frac{c_2}{c_1} - \beta_2\right)^+$$

If exploration is radical, then

$$w_1 = \frac{c_1(1+E[1_{R_2>B_1}])op_{21}}{E\left[1_{R_2>B_1}\left(\overline{op_{22}^{12}}-op_{21}\right)\right]}\left(\frac{c_2}{c_1} - \beta_2\right)^+$$

And

$$\lambda_2 = \frac{c_1}{op_{21}-op_{20}} - \frac{(1+E[1_{R_2>B_1}])c_1}{E\left[1_{R_2>B_1}\overline{op_{22}^{12}}\right]-p_0 op_{20}}\left(\beta_1 - \frac{c_2}{c_1}\right)^+ + \frac{(1+E[1_{R_2>B_1}])c_1}{E\left[1_{R_2>B_1}\overline{op_{22}^{12}}\right]-p_0 op_{21}}\left(\frac{c_2}{c_1} - \beta_1\right)^+ + \left(\frac{c_2}{c_1} - \beta_2\right)^+ * \frac{(1+E[1_{R_2>B_1}])(E[1_{R_2>B_1}]-p_0)op_{21}c_1}{E\left[1_{R_2>B_1}\left(\overline{op_{22}^{12}}-op_{21}\right)\right]\left(E\left[1_{R_2>B_1}\overline{op_{22}^{12}}\right]-p_0 op_{21}\right)}$$

To implement exploration, the principal must prevent the agent from shirking or exploitation. The principal does not make any payments to the agent after a failure in the second period, since this only gives an incentive for the agent to shirk. Moreover, the principal does not make payments to the agent after a success in the first period for two reasons. First, rewarding first-period success gives the agent the incentive to employ the conventional work method in the first period, since the initial expected probability $E[p_2]$ of success with the new work method is lower than the probability p_1 of success using the conventional work method. Second, when there is success in the first period, additional information about the first-period action is provided by the second-period performance, since the expected probability of success with the new work method in the second period depends on the action taken by the agent in the first period. Delaying compensation to obtain this additional information is, therefore, optimal.

The principal expects the agent to choose the conventional work method in the second period after a failure in the first period. To prevent the agent from shirking in this situation, the principal pays the agent $\lambda_3 = \frac{c_1}{op_{21}-op_{20}}$.

Then, finally, to encourage exploration, the principal must reward the agent's second-period success following a success in the first period. The wage rate λ_2 depends on the difficulty of implementing the exploration relative to exploitation. With the increase in c_2/c_1, the difficulty of implementing exploration relative to exploitation increases, and the wage rate λ_2 must increase, too.

If $c_2/c_1 < \beta_1$, then exploitation is too costly for the agent, but exploration is not costly for the agent. In this situation, the principal pays the agent $\lambda_2 < \lambda_3$. If $c_2/c_1 \geq \beta_1$, then exploitation is not too costly for the agent, but exploration is costly for the agent. In this situation, the principal must pay the agent $\lambda_2 \geq \lambda_3$. When $c_2/c_1 \geq \beta_2$, the rage rate λ_2 must increase further. In this case, if $\frac{E[1_{R_2 \leq B_1}]}{E[1_{R_1 \leq B_1}]} \geq \frac{E\left[1_{R_2>B_1}\overline{op_{22}^{12}}\right]}{E[1_{R_1>B_1}]op_{21}}$, namely exploration, is radical, it has a high expected probability of failure in the first period relative to the probability of failure of the conventional action. The expected reward for exploration of two consecutive successes cannot compensate for the risk of failure. So, the principal must pay the agent a higher λ_2, and reward the agent for failure in the first period at the same time.

Similarly, baseline B_t will affect the results. If $B_2 \geq B_1$, then λ_3 and λ_2 increase. This can be interpreted as when the baseline of the standard for success increases, the difficulty for success in the second period increases and the exploitation constraint may be binding. To prevent exploitation, the principal must pay the agent an extra premium if there is success in the second period.

To illustrate the differences in the optimal contracts between these two action plans, see figure 2 as follows.

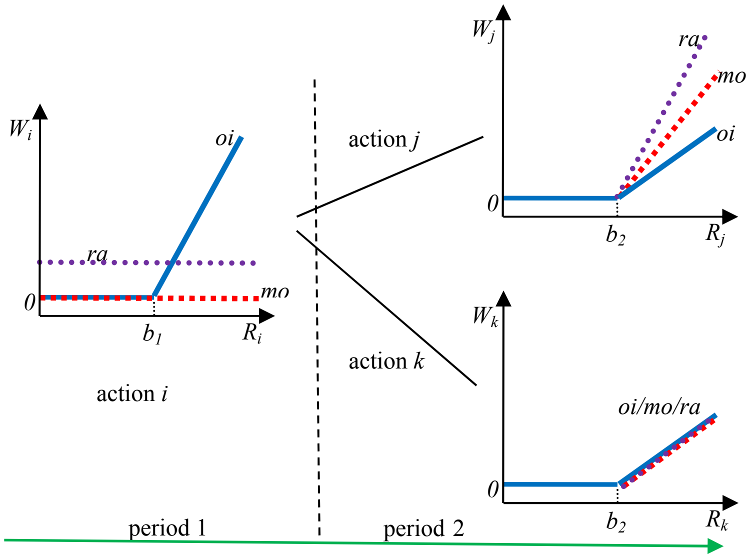

Figure 2: Structured Reward of Action Plan $\langle i_k^j \rangle$

In figure 2, the reward W received by the agent for different action plans are displayed. The blue solid line is for the action plan $\langle 1_1^1 \rangle$ (exploitation, simplified as oi), the red dashed line for action plan $\langle 2_1^2 \rangle$ (moderate exploration, simplified as mo) and the purple dotted line for action plan $\langle 2_1^2 \rangle$ (radical exploration, simplified as ra). The sign *oi/mo/ra* means the three actions of exploitation and exploration (moderate and radical) have the same wage and wage rate in this situation, These three lines coincide with each other. Given the optimal contracts, reward $W\langle i_k^j \rangle$ is dependent on the output $R\langle i_k^j \rangle$ of the action in every period. We can see that when the principal provides the incentive structure as $\lambda_1 \geq \lambda_2 = \lambda_3$, the agent will usually take action plan $\langle 1_1^1 \rangle$ (exploitation) because, in the first period, action 1 can produce most output with the highest probability and he can earn the most rewards. In the second period, whatever action he takes, he will get the same wage rate. In this situation, he will continue to take action 1 because there is no new information about action 2 without the tests from the first period. Therefore, the optimal choice is still action 1. However, if the principal provides the incentive structures as per action plan $\langle 2_1^2 \rangle$, the agent will use the exploration action plan. In the first period, the rewards do not depend on output R_i, but in the second period, the rewards are highly dependent on the output. If he takes action 2 in the first period and gets information about it, he can get more rewards in the second period. If he has a success in the first period with action 2, he will continue to take action 2 in the second period, and get the highest rewards. If he has a failed experience in the first period with action 2, he can turn to action 1 in the second period, and get the same rewards as action $\langle 1_1^1 \rangle$. If the possibility of failure with action 2 is higher, he will get a fixed reward in the first period as compensation. The optimal results in our models show that the rewards of the agent depend not only on the output, but also on the path of the performance of his output.

The optimal contract results for propositions 1 and 2 have several implications in the real world. They can explain many things in relation to managerial compensation, such as a combination of stock options with long vesting periods, and option re-pricing. Stock options can be presented to the managerial staff, and even to the ordinary people in the companies. This is one kind of incentive method that can solve the problem of the Principal-agent, or be used to motivate ordinary people to exert themselves to earn more rewards in the long term. This compensation policy is more fashionable in high-tech companies or venture capital projects, where innovation is their basic property. They generally required that stock options must be with a long vesting period, which means that the holder of stock options cannot sell the options in the short term. Here, the optimal contracts of our model present a high non-fixed wage rate for the second period that motivates the agent to undertake exploration. In the process of solving the optimization problem, it shows that the rewards the agent can earn are the product of non-fixed wage rates and options. This is consistent with the long vesting period stock options. According to the re-pricing options, because the innovation process is split into several periods in the real world, the innovation path may be changed over time because the conditions and circumstances may change. Therefore, when gathering the optimal results of the innovation practices, the structured reward contracts must be adjusted over time, as this is the re-pricing of options.

The results produced here in our models can be tested in the empirical world. One can test whether the incentive contracts are used in exploration practices; and whether these incentive contracts have an effect and lead to more innovation. Furthermore, as these incentive contracts are structured or similar to the optimal contracts in our model, the question arises whether these structured contracts lead to further innovation. If not, which one is the most suitable?

We are able to undertake some empirical work with Chinese companies, especially high-tech companies. Currently in China, innovations are very important, from the whole country to a single company. How to motivate for innovation is the key to this trend. Using our results, we can undertake empirical work to find the characteristics of innovation practices and revise our theory results further. Additionally, these results will be of help to our innovative country of China. This work will be done in the near future.

5. Conclusion and Limitations

Based on the framework of Manso (2011), this paper has studied the incentives for innovation with non-fixed rewards for the agent. We have explored the standard of success, and the reward of the agent depends on the amount of the excess output over the baseline. The fixed wage and wage rate for success have been designed. These structured incentives can motivate the agent to select a more innovative work method and encourage them to exert effort to obtain a better output.

The optimal contract that implements both exploitation and exploration is comprised of a series of options, which are structured. To stimulate exploration, the principal must offer a proper fixed reward so as to tolerate the possibility of failure; at the same time, a non-fixed reward must not be offered. The optimal contract depends on the baseline of success and the private costs of the agent, especially the cost ratio of exploration and exploitation.

There are some limitations to the paper.

(1) We have only considered the first-order stochastic dominant relationship between the returns. They may be either second-order or higher-order. Therefore, additional real distributions are needed to discuss the problem further.

(2) In the paper, the information has assumed symmetry. In fact, the information may be asymmetrical, which will impact on the results severely.

(3) The interest rate and time preferences are not considered. The span of the periods may have an important impact on the solutions.

(4) Here, we have only considered problems over two periods. Although two periods can demonstrate the basic insights of the problem, new results

may be found over more periods, especially for infinite periods. In fact, whether the innovation test can be obtained continuously is a problem. For example, the termination will be a threat for the agent. Additionally, in the long run, if the exploration has been tested for many times and has success, will it return to a conventional action, namely, the exploratory action turns into an exploitation action, and rewards must be changed accordingly. All these situations must be considered in our next study.

(5) However, some of the predictions of this model remain untested, and additional empirical work is required. Because there are so many moderators in the real world, the basic optimal results, in theory, may not be optimal. Therefore, again, more empirical work must be done. We are going to be doing this in the next step. We will collect enough data to test our model, or look for the properties of innovation motivation in the real world, and revise our theoretical model accordingly. Based on this empirical work, we expect to ascertain several parameters related to this problem. We can then run some simulations for a variety of situations.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (NSFC), Grant No.71171095 and the Chinese Scholarship Council (grant no.201308420229).

References

Aghion, P., & Tirole, J. (1994). The management of innovation. *The Quarterly Journal of Economics, 109*(4), 1185-1209.

Arrow, K. (1969). Classificatory notes on the production and diffusion of knowledge. *The American Economic Review* 59, 29-35.

Berle, A., & Means, G. (1932). *The Modern Corporation and Private Property*. New York, NY: Macmillan.

Ederer, F., & Manso, G. (2013). Is pay-for-performance detrimental to innovation? *Management Science, 59*(7), 1496 - 1513.

Harris, M., & Raviv, A. (1978). Some results on incentive contracts with applications to education, insurance, and law enforcement. *The American Economic Review, 68*(1), 20-30.

Hellmann, T., & Veikko Thiele, V. (2011). Incentives and innovation inside firms: A multi-tasking approach. *American Economic Journal: Microeconomics, 3*(1), 78-128.

Holmstrom, B. (1979). Moral hazard and observability. *The Bell Journal of Economics, 10*(1), 74-91.

Holmstrom, B. (1989). Agency costs and innovation. *Journal of Economic Behavior and Organization, 12*(3), 305-327.

Manso, G. (2011). Motivating Innovation. *The Journal of Finance, 66*(5), 1823-1860.

March, J. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2*(1), 71-87.

Moscarini, G., & Smith, L. (2001). The optimal level of experimentation. *Econometrica, 69*(9), 1629-1644.

Tian, X., & Wang, T. (2014). Tolerance for failure and corporate innovation. *The Review of Financial Studies, 27*(1), 211-255.

Zheng, Chengli & Chen, Y. (2013). Motivating innovation with a structured incentives scheme – Extending the Manso model. *African Journal of Business Management, 7*(37), 3743-3753.

About Authors

Chengli Zheng is a full Professor from School of Economics and Business Administration, Huazhong Normal University, China. And he is a visiting Scholar in Department of Applied Mathematics & Statistics, Stony Brook University, USA. He is active in financial economics and management. He has published over 30 articles in journals the related fields.

Yan Chen is a lecturer in School of Mathematics & Statistics, Huazhong Normal University, China. Her main interest involves financial accounting and management. She has published over 10 articles in journals of the related fields.