The Pygmalion and Galatea Effects: An Agency Model with Reference-Dependent Preferences and Applications to Self-Fulfilling Prophecy*

Kohei Daido† Hideshi Itoh‡

First Version: February 18, 2005
This Version: July 31, 2006

Abstract

We attempt to formulate and explain two types of self-fulfilling prophecy, called the Pygmalion effect (if a supervisor thinks her subordinates will succeed, they are more likely to succeed) and the Galatea effect (if a person thinks he will succeed, he is more likely to succeed). To this purpose, we extend a simple agency model with moral hazard and limited liability by introducing a model of reference-dependent preferences (RDP) by Kőszegi and Rabin (2004). We show that the agent with high expectations about his performance can be induced to choose high effort with low-powered incentives. We then argue that the principal's expectation has an important role as an equilibrium selection device, when the level of the agent's ability is intermediate.

JEL Classification Numbers: B49, D82, M12, M52, M54

Keywords: Self-fulfilling prophecy, Pygmalion effect, Galatea effect, reference-dependent preferences, agency model, moral hazard

*Earlier versions were circulated under the title "The Pygmalion Effect: An Agency Model with Reference Dependent Preferences." We are grateful to Kenichi Amaya, Munetomo Ando, Björn Bartling, Leonardo Felli, Michihiro Kandori, and seminar participants at CESifo Area Conference on Applied Microeconomics, ChengChi University, Kwansei Gakuin University, Osaka University, and the 20th Annual Congress of the European Economic Association, for helpful comments. Financial support from the 21st Century COE program "Dynamics of Knowledge, Corporate System and Innovation" at Graduate School of Commerce and Management, Hitotsubashi University, and Murata Science Foundation is gratefully acknowledged.

†School of Economics, Kwansei Gakuin University.

‡Graduate School of Commerce and Management, Hitotsubashi University.

> I shall always be a flower girl to Professor Higgins because he always treats me as a flower girl and always will; but I know I can be a lady to you because you always treat me as a lady and always will.
>
> —Eliza Doolittle, in *Pygmalion* by George Bernard Shaw

1 Introduction

People (pupils, subordinates, and so on) tend to act in accordance with the expectation of others (teachers, managers, and so on). In particular, the former may, to some degree, internalize the higher expectations placed on them by the latter, and then act in ways to fulfill those expectations. A pioneering work by Rosenthal and Jacobson (1968) shows, through their experimental research, that a teacher's expectation for a pupil's intellectual competence can come to serve as an educational self-fulfilling prophecy, and names this phenomenon the *Pygmalion effect* after Greek myths.[1] Livingston (1969) discusses the Pygmalion effect in managerial setting.[2] He argues that a number of case studies and experiments reveal the following:

(a) "What managers expect of subordinates and the way they treat them largely determine their performance and career progress."

(b) "A unique characteristic of superior managers is the ability to create high performance expectations that subordinates fulfill."

(c) "Less effective managers fail to develop similar expectations, and as a consequence, the productivity of their subordinates suffers."

[1] A series of research by Rosenthal and his collaborators studies the Pygmalion effect in educational setting. Jussim (1986) provides a theoretical model of the Pygmalion effect in the classroom.

[2] See also Goddard (1985).

(d) "Subordinates, more often than not, appear to do what they believe they are expected to do."

Since Livingston (1969), many researchers have been studying the Pygmalion effect in business or military organizations. Kierein and Gold (2000) and McNatt (2000) conduct meta-analysis of relevant studies within management contexts, and both find that the Pygmalion effect is in general fairly strong.

From these existing studies, we can summarize the way the Pygmalion effect occurs as follows:

1. A manager's high expectation influences her attitude toward her subordinates.

2. Such attitude has positive effects on subordinates' self-expectancy.

3. The subordinates' enhanced self-expectancy then improves their performance.

The last part of this process, that a person's enhanced self-expectation improves his own performance, is often called the *Galatea effect*. For example, Kierein and Gold (2000) explain the Galatea effect as one of other types of expectation effects: "The Galatea effect occurs not when the leader has expectations of subordinates, but when subordinates' raised expectations of themselves are realized in their higher performance." They however state that it is part of the Pygmalion effect, and examine the Pygmalion and Galatea effects together in their meta-analysis.[3]

In this paper we attempt to formalize and explain both the Pygmalion and the Galatea effects. To this purpose, we extend a simple but standard

[3]On the other hand, McNatt (2000) seems to emphasize that the feature that a manager's expectation has the impact on her subordinate's self-expectancy appears uniquely in the Pygmalion effect.

model of a principal and an agent with moral hazard and limited liability. A key extension from this standard model is that the agent has *reference-dependent preferences* (henceforth RDP). What a person has RDP means that his preferences are conditional on a reference point, and various anomalies such as loss aversion, endowment effects, status quo bias, and so on, are consistent with RDP.[4] More precisely, the payoff depends on the realized consumption as gain or loss relative to a reference level. What serves as the reference point is thus crucial to the model with RDP. In this respect, Masatlioglu and Ok (2005) and Sagi (2004) as well as most experimental research assume that the status quo serves as the reference point, and Sugden (2003) considers the reference point as one's current endowment which is determined by a "reference lottery." Note that in these studies reference points are exogenously given.

Our model is built on a yet another model of RDP by Kőszegi and Rabin (2004), which has the following three important features.[5] First, it combines the standard consumption payoff that is independent of the reference level, with the reference-dependent gain-loss payoff that depends solely and in a "universal" way on consumption payoff relative to the reference level. Second, a person's reference point is her *recent expectation*, represented by a probability measure, over outcomes.[6] Third, the model by Kőszegi and Rabin (2004) has a prominent feature that the reference point is endogenously determined by the person's rational expectation.[7] To this end, they define the *personal equilibrium* which requires that a person maximize his payoff

[4]The seminal paper by Kahneman and Tversky (1979) explores issues on RDP. Recently, various models of RDP have been developed. For example, Masatlioglu and Ok (2005), Sagi (2004), and Sugden (2003) give axiomatic foundations for models of RDP.

[5]We give a brief sketch of their model in the next section.

[6]In this sense, their model is similar to that of Sugden (2003).

[7]Munro and Sugden (2003) and Falk and Knell (2004) also study the endogenous determination of reference points.

given his rational expectations about outcomes, and hence the expectations themselves depend on his own anticipated behavior. These features of their model allow us to introduce the reference-dependent nature of human decisions into standard economic problems, a principal-agent problem in this paper.[8]

Following their model, we suppose that the agent's utility depends not only on his material payoff, as in the standard model, but also on the gain-loss payoff which is defined by the agent's evaluation of his consumption bundle as gains or losses relative to a reference point. The reference point of the agent is his expectation about the effort level chosen by the agent and the resulting success probability of the project. We then define the personal equilibrium that determines the agent's reference point endogenously. Our main contribution is to analyze interaction between RDP and incentives designed by the principal. We first take a contract as given, and analyze the agent's personal equilibrium. We show that compared with the standard model without RDP, the agent's higher expectation enables the principal to implement high effort with lower-powered incentives. We also show that when the power of incentives is intermediate, multiple equilibria may exist. In this case the agent's expectations are self-fulfilling: he chooses high effort if he expects to do so, while he chooses low effort if it is his expected effort.

We then study the optimal contract solving the principal's problem. The principal's contract affects the agent's personal equilibrium and hence his expectations. Furthermore, the principal wants to make the agent attend to high effort in the region with multiple equilibria. Analysis of optimal contracts depends on how the agent forms his expectation in this region.

[8]Our work is thus in the spirit of one of the goals of their paper. "In addition to the ways we believe our model substantially clarifies and improves on existing theories of reference-dependent preferences, it also has an attractive "methodological" feature in promoting the addition of reference dependence to existing economic models. (p.6)"

We thus explicitly distinguish among the following three cases: (i) the principal chooses the personal equilibrium she prefers; (ii) the agent chooses the personal equilibrium he prefers; and (iii) the agent forms his expectation consistent to his inherent type where we call the agent with a higher expectation *optimistic* type and one with a lower expectation *pessimistic* type. We mainly study the third case by analyzing various contractual arrangements that are different concerning how the principal deals with the pessimistic type.

Based on the analysis of our principal-agent model, we formalize the Pygmalion and Galatea effects. Since the literature on these effects mainly focuses on effects of the subjects' abilities under given incentive schemes, we restate our results in terms of a measure of the agent's ability, which is observable to both the principal and the agent. We show that high effort is a personal equilibrium if the agent's ability is sufficiently high. Thus if the agent expects he can succeed with sufficiently high probability, then he actually succeeds with the same high probability. We interpret this as the Galatea effect: The agent's self-expectation about his performance determines his actual performance. Concerning the Pygmalion effect, the principal does not play a role if the agent's ability is higher than a threshold level, or lower than another threshold level. However, in an intermediate range of abilities, there are multiple equilibria, and we interpret the principal's expectation as an equilibrium selection device: the Pygmalion effect realizes as the principal's attempt to make the agent choose a particular equilibrium is successful.[9]

[9]If the Pygmalion effect works, then the principal prefers the agent's decisions to reflect his loss aversion, in contrast to the following statement of Tversky and Kahneman (1991): "the principal may not wish the agent's decisions to reflect the agent's aversion to losses, because the agent's reference level has no bearing on the principal's experience of outcomes. (p.1058)"

Although we are unaware of any economic literature studying the Pygmalion effect, there are possible alternative approaches to explaining this effect. First, there is a simple explanation that the principal with high expectation takes some explicit actions to improve the agent's productivity. This is nothing but the theory of human capital. In contrast to this explanation, we consider the situation where the principal's expectation implicitly influences the agent's performance. In other words, our model does not have any component which directly affects the agent's productivity. Future experimental work should test whether or not a mere expectation of high performance is fulfilled.

The second, more interesting approach is to focus on the role of information transmission by the principal. We do not consider the issue on asymmetric information in this paper, by assuming that information concerning the agent's ability is symmetric. However, the principal with high expectation may effectively transmit her private information on the agent's productivity when the agent does not know his own productivity.[10] Bénabou and Tirole (2003) is an example along this line.[11] In their model, the principal's policy (wage scheme) as a signal informs the agent of his ability and then affects his action: costly signal from the principal serves as a motivational device for the agent. However, we think this kind of explanation is not a whole story for the following two respects. First, the Pygmalion effect works even when agents know their own abilities without such informative signals. An experimental result illustrated in Livingston (1969) shows this point. At an office of the Metropolitan Life Insurance Company, a manager decided to group his insurance agents according to their abilities,

[10]Note that the principal's mere expression of her expectation is not credible to the agent, since it is just cheap talk.

[11]See also Hermalin (1998).

which fact was known symmetrically among agents. Then, the performance of the agents in the top group improved dramatically, and that of those in the lowest unit declined. However, the performance of those in the average group also improved significantly, due to the supervisor's high expectation. This observation is consistent with our results. Second, Bénabou and Tirole (2003) assume that the agent devotes more effort when he receives a good signal that convinces him that his ability is high. That is, the Galatea effect is exogenously given. In our model, in contrast to this signaling approach, there is no information transmission, and both the Galatea and the Pygmalion effects are explained endogenously.

The rest of the paper is organized as follows. In section 2 we explain the model of RDP based on Kőszegi and Rabin (2004). In section 3 we build a simple agency model with RDP, and analyze the personal equilibrium and the optimal incentive scheme in section 4. In section 5 we relate our results with the Pygmalion and Galatea effects. Section 6 concludes.

2 Reference-Dependent Preferences

Following Kőszegi and Rabin (2004), we formulate the reference-dependent nature of preferences in the following way. Let $c = (c_1, \ldots, c_n)$ be a consumption bundle of an agent, and $r = (r_1, \ldots, r_n)$ be a reference consumption bundle. How reference point r is determined will be explained in the next section. We define the agent's overall payoff $u(c \mid r)$ by

$$u(c \mid r) = v(c) + z(c \mid r) = \sum_{k=1}^{n} v_k(c_k) + \sum_{k=1}^{n} z_k(c_k \mid r_k), \tag{1}$$

where $v(c)$ is his material payoff, as in standard models, and $z(c \mid r)$ represents the agent's evaluation of his consumption bundle as gain and loss relative to a reference point. We call this part of the agent's payoff as gain-

loss payoff. (1) implies that each dimension of consumption is assumed to be additively separable.

The model is extended to cases in which there is uncertainty in consumption outcomes as well as reference points. Let F be the probability distribution function of consumption bundle c, and G be the distribution function of reference point r. The agent's payoff is then given by

$$U(F \mid G) = \int_c \int_r u(c \mid r) dF(c) dG(r).$$

This formulation means that the agent compares a given outcome c with all r in the support of the reference lottery.[12] For example, suppose that the reference lottery is between $-\$100$ and $\$200$ with equal probability. If the actual outcome is $\$0$, then he expects to feel a gain relative to $-\$100$ and a loss relative to $\$200$, with equal probability.

As Kőszegi and Rabin (2004) do, we further assume that each dimension is evaluated by the same "universal" gain-loss function $\mu(\cdot)$ of the difference of consumption from the reference level, evaluated by material payoff:

$$z(c \mid r) = \sum_{k=1}^{n} \mu(v_k(c_k) - v_k(r_k)) \tag{2}$$

The gain-loss function $\mu(\cdot)$ is assumed to have the following properties. They capture important features of how people evaluate gain and loss from the reference point.

A0 $\mu(0) = 0$ and $\mu'(y) > 0$.

A1 $\mu''(y) \leq 0$ for $y > 0$, and $\mu'(y) > 0$ and $\mu''(y) \geq 0$ for $y < 0$.

A2 If $y > y' > 0$, $\mu(y) - \mu(y') < \mu(-y') - \mu(-y)$ holds.

A3 $\lim_{y \uparrow 0} \mu'(y) / \lim_{y \downarrow 0} \mu'(y) \equiv \lambda > 1$.

[12] In Sugden (2003), an outcome is compared only to the outcome that would have resulted from the reference lottery in the same state.

A1 represents *diminishing sensitivity*, implying that as the consumption level moves further away from the reference level, the marginal valuation of gains and losses decreases. And A2 and A3 capture *loss aversion*, A2 for "large" stakes and A3 for marginal ones.

In what follows we assume away diminishing sensitivity and isolate the effect of loss aversion, by assuming $\mu(\cdot)$ is linear, and define

$$\mu(v_k(c_k) - v_k(r_k)) = \begin{cases} \alpha(v_k(c_k) - v_k(r_k)) & \text{if } v_k(c_k) - v_k(r_k) > 0, \\ \alpha\lambda(v_k(c_k) - v_k(r_k)) & \text{if } v_k(c_k) - v_k(r_k) < 0, \end{cases} \tag{3}$$

where α, a positive constant, is the weight on the gain-loss payoff, and $\lambda > 1$ is the "coefficient of loss aversion" which is the same as λ defined in A3.

3 A Simple Agency Model with RDP

There are two risk neutral parties, a principal and an agent. The agent engages in one project on behalf of the principal. The outcome of the project is either success (s) or failure (f), and the probability distribution depends on the agent's effort. We assume there are two feasible effort levels e_0 and e_1, and denote by p_i the probability of success under effort e_i.[13] We assume $0 < p_0 < p_1 < 1$ and denote $\Delta_p \equiv p_1 - p_0$.

In the standard agency model, the agent's payoff depends on his "consumption" bundle (w, e_i) where w is remuneration received from the principal. Let $v(w, e_i)$ be his *material* payoff function, and assume it is additively separable: $v(w, e_i) = w - d_i$ where d_i is the agent's private cost of effort e_i. For simplicity we assume $d_0 = 0 < d_1$, and denote $d = d_1$. Using the formulation introduced in section 2, we extend this standard model as follows. Let $(\overline{w}, e_j)$ be a reference point, and define the agent's overall payoff $u(w, e_i \mid \overline{w}, e_j)$ by

$$u(w, e_i \mid \overline{w}, e_j) = w - d_i + \mu(w - \overline{w}) + \mu(d_j - d_i) \tag{4}$$

[13] The analysis can be extended to the case where the effort variable is continuous.

where the gain-loss function $\mu(\cdot)$ is defined by (3).

The agent's effort is unobservable to the principal, while the outcome of the project is verifiable. The principal can thus design an incentive compensation scheme (b_s, b_f) where b_i is remuneration paid from the principal to the agent when outcome is $i \in \{s, f\}$. We assume that b_i must satisfy the limited liability constraint $b_i \geq 0$. We also denote the difference in payment by $\Delta_b = b_s - b_f$.

The timing of the game is as follows.

1. The principal offers a contract.
2. The agent either accepts or rejects the contract. If he rejects it, the game ends and each of the parties receives the reservation payoff zero. If the agent accepts the contract, the game moves to the next stage.
3. The reference point of the agent is determined.[14]
4. The agent chooses effort.
5. The outcome of the project realizes and the payment is made according to the contract.

We now discuss how the agent's reference point is determined. First, we follow Kőszegi and Rabin (2004) by taking the standpoint that the reference point is determined by the agent's recent expectations about what he is going to get. In most literature, the reference point of an individual is given exogenously as his current or past endowments, while little is known both theoretically and empirically concerning how the reference point is determined. Kőszegi and Rabin (2004) argue that expectations play a central role in determining reference points. For example, they state as follows.

[14] Our results are not affected by an alternative timing as long as the agent determines his reference point *after* the principal offers a contract and *before* he chooses effort.

"While existing experimental evidence is generally interpreted in terms of the reference point being the endowment or status quo, we feel that virtually all of this evidence can also be interpreted in terms of expectations—for the simple reason that in the contexts studied people plausibly expect to keep the status quo. (p.3)" Expectations seem to play a central role in determining reference points even when endowments do not. For example, workers are averse to wage cut not because it reduces their status quo level of wealth but because they expect to receive lower wages. People tend to feel a loss if they expect to buy a good and find it is sold out at the shop. Since we study the effects of expectations on performance, their formulation in particular fits well with our research agenda.

If we adopt the expectation-as-reference view and apply it to our agent, the agent's preferences depend on his expectations, which themselves depend on his preferences. The agent with some predictive ability will take this feedback into account, and reach a state in which his expectations are consistent with his eventual outcomes. Following Kőszegi and Rabin (2004), we thus model the agent's decision making in terms of an "equilibrium" as follows.

Suppose that a compensation scheme (b_s, b_f) has been accepted by the agent. The agent's reference point consists of effort e_j he is expected to choose, and the resulting probability distribution over (b_s, b_f), which is represented by the probability of success p_j.[15] Then (e_j, p_j) is a *personal equilibrium* if for $i \neq j$,

$$U(e_j, p_j \mid e_j, p_j) \geq U(e_i, p_i \mid e_j, p_j) \tag{5}$$

[15]Since p_j is uniquely determined by e_j, it is enough to define reference points and equilibria in terms of effort only. We however include the probability distribution in order to emphasize the existence of gains and losses in terms of outcome-dependent payments.

where $U(\cdot)$ is the agent's expected payoff and is given by

$$\begin{aligned} U(e_i, p_i \mid e_j, p_j) = b_f + p_i\Delta_b - d_i \\ + p_i(1-p_j)\mu(\Delta_b) + (1-p_i)p_j\mu(-\Delta_b) \\ + \mu(d_j - d_i). \end{aligned} \tag{6}$$

The first line of (6) is the expected material payoff. The second and third lines represent the gain-loss payoff under reference point (e_j, p_j). For example, suppose that the reference point is (e_1, p_1), and the agent's choice is (e_0, p_0). The agent enjoys gain αd (corresponding to the third line) because he saves cost d by not choosing e_1 which he was expecting to choose by spending cost d. If the agent succeeds, he enjoys gain $\alpha\Delta_b$ with probability $1 - p_1$ because he was expecting to fail with this probability. Similarly, when the agent actually fails, he suffers from loss $\alpha\lambda\Delta_b$ with probability p_1 since he was expecting to succeed with this probability. These gain and loss correspond to the second line.

Note that the crucial feature behind the agent's preferences is that there is time lag between the stage when the agent forms expectations (stage 3 in the timing of the game) and the stage when the agent chooses effort (stage 4) or when uncertainty is resolved (stage 5): The agent's preferences do not change immediately at stage 4 or 5, and hence he compares his effort and income with what he expected them to be at stage 3.

The definition of the personal equilibrium states that if the agent's reference point is the expectation to choose e_j and hence to succeed with probability p_j, then he should indeed be willing to choose e_j. The reference point is thus determined endogenously by the agent, anticipating his choice, and then given the reference point, the agent chooses effort consistent with his expectation.

4 Analysis

In this section, we study the agent's personal equilibrium and the optimal contract. In subsection 4.1, we first find the personal equilibrium and then examine its characteristics. In subsection 4.2, we derive the optimal contract.

4.1 The Agent's Personal Equilibrium

Suppose that the principal offers a contract (b_s, b_f), which is accepted by the agent. Given the contract, we analyze the agent's personal equilibrium. In the next subsection, we analyze the optimal contract.

There are two candidates for personal equilibria, (e_1, p_1) and (e_0, p_0).[16] First consider (e_1, p_1). The relevant expected payoffs are calculated as follows:

$$U(e_1, p_1 \mid e_1, p_1) = b_f + p_1\Delta_b - d + p_1(1-p_1)\alpha\Delta_b - (1-p_1)p_1\alpha\lambda\Delta_b$$

$$U(e_0, p_0 \mid e_1, p_1) = b_f + p_0\Delta_b + \alpha d + p_0(1-p_1)\alpha\Delta_b - (1-p_0)p_1\alpha\lambda\Delta_b$$

Pair (e_1, p_1) is a personal equilibrium if $U(e_1, p_1 \mid e_1, p_1) \geq U(e_0, p_0 \mid e_1, p_1)$, or

$$\Delta_b \geq \beta_1 \equiv \frac{d}{\Delta_p}\frac{1+\alpha}{1+\alpha+\alpha p_1(\lambda-1)} \tag{PE1}$$

Next consider (e_0, p_0). The following expected payoffs are relevant.

$$U(e_0, p_0 \mid e_0, p_0) = b_f + p_0\Delta_b + p_0(1-p_0)\alpha\Delta_b - (1-p_0)p_0\alpha\lambda\Delta_b$$

$$U(e_1, p_1 \mid e_0, p_0) = b_f + p_1\Delta_b - d - \alpha\lambda d + p_1(1-p_0)\alpha\Delta_b - (1-p_1)p_0\alpha\lambda\Delta_b$$

[16] A mixed-strategy equilibrium also exists when both (e_1, p_1) and (e_0, p_0) are equilibria. However, the agent's expected payoff under the mixed-strategy equilibrium is lower than the expected payoff under one of two pure-strategy equilibria, and hence we focus on pure strategies.

Then (e_0, p_0) is a personal equilibrium if $U(e_0, p_0 \mid e_0, p_0) \geq U(e_1, p_1 \mid e_0, p_0)$, or

$$\Delta_b \leq \beta_0 \equiv \frac{d}{\Delta_p} \frac{1 + \alpha\lambda}{1 + \alpha + \alpha p_0(\lambda - 1)} \tag{PE0}$$

It is easy to show the following results.

Proposition 1. (i) $\beta_0 > d/\Delta_p > \beta_1$. (ii) When contract (b_s, b_f) is given, there are three ranges of "incentive intensity" Δ_b that characterize personal equilibria.

(a) If $\Delta_b > \beta_0$, then (e_1, p_1) is the only personal equilibrium.

(b) If $\Delta_b < \beta_1$, then (e_0, p_0) is the only personal equilibrium.

(c) If $\beta_1 \leq \Delta_b \leq \beta_0$, both (e_0, p_0) and (e_1, p_1) are personal equilibria.

To understand the results, consider first the standard agency model in which the agent does not exhibit RDP (corresponding to $\alpha = 0$ in our model). The agent then prefers to choose e_1 if $\Delta_b > d/\Delta_p$, and choose e_0 if $\Delta_b < d/\Delta_p$. Proposition 1 (i) then implies that the agent with RDP, when he expects to choose e_1, actually chooses e_1 for incentive intensity Δ_b lower than d/Δ_p, the critical value under the standard case. And the agent, expecting to choose e_0, actually chooses e_0 for Δ_b higher than d/Δ_p. In other words, there are additional incentives from the agent's reference dependence or more specifically, loss aversion.

Intuitively, Proposition 1 (i) can be understood as follows. First consider result $d/\Delta_p > \beta_1$. RDP introduce the following positive incentive effects. If the agent, expecting the outcome to be success with probability p_1,chooses e_1 instead of e_0, then the chance of "gain" increases from $p_0(1 - p_1)$ to $p_1(1 - p_1)$, and that of "loss" decreases from $(1 - p_0)p_1$ to $(1 - p_1)p_1$. These effects reinforce the incentive to choose e_1 via incentive pay. On the other

hand, there is a negative effect from RDP. When the agent's reference effort is e_1, shirking (e_0) benefits the agent by saving the cost of effort. This effect is represented by αd. The positive effects dominate because of $\lambda > 1$, the agent's loss aversion.

The intuition behind $\beta_0 > d/\Delta_p$ is similar. When the agent's reference point is (e_0, p_0), e_1 is more attractive under RDP than without RDP, because gain (loss) is more (less, respectively) likely. However, the agent is more reluctant to choose e_1 because he experiences loss $\alpha\lambda d$. This latter negative effect dominates because of loss aversion, and thereby the agent chooses e_0 even though incentives strong enough to induce e_1 are provided for the agent without RDP.

Figure 1 illustrates three ranges of Δ_b in Proposition 1 (ii). In regions $\Delta_b > \beta_0$ and $\Delta_b < \beta_1$, the agent's preferences exhibit a status quo bias, in the sense that if the agent with a reference point (e_j, p_j) wants to select (e_i, p_i), he prefers (e_i, p_i) when it is his reference point.[17] In the intermediate range, there are multiple personal equilibria. In this range, the agent expecting to choose e_1 actually do so in order to reduce the probability of loss from the failure of the project, while he chooses e_0 to avoid loss due to unexpected disutility of effort, if e_0 is his expected effort.

The next proposition reports some comparative statics results.

Proposition 2.

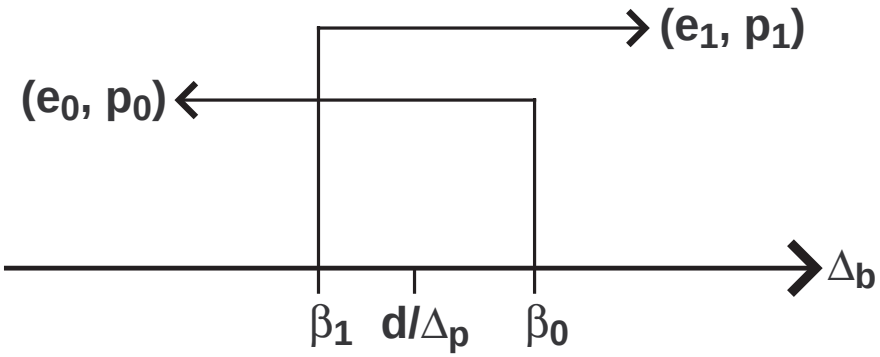
(a) β_0 is increasing in α and λ, and decreasing in p_1.

(b) β_1 is decreasing in α, λ and p_1.

(c) $\beta_0 - \beta_1$ is increasing in α, λ and decreasing in p_1.

[17] This result is a special case of Proposition 1 in Kőszegi and Rabin (2004).

Figure 1: Personal Equilibria and Incentive Intensity



Proof. Results (a) and (b) directly follow from the definitions in (PE0) and (PE1). For (c), the comparative statics results for α and λ are immediate from (a) and (b). By differentiating β_0 and β_1 by p_1, we obtain

$$\begin{aligned}
\frac{\partial \beta_0}{\partial p_1} - \frac{\partial \beta_1}{\partial p_1} &= -\frac{d}{\Delta_p^2}\left(\frac{1+\alpha\lambda}{1+\alpha+\alpha p_0(\lambda-1)} - \frac{1+\alpha}{1+\alpha+\alpha p_1(\lambda-1)}\right) \\
&\quad + \frac{d}{\Delta_p}\frac{\alpha(\lambda-1)(1+\alpha)}{(1+\alpha+\alpha p_1(\lambda-1))^2} \\
&\leq -\frac{d}{\Delta_p^2}\frac{\alpha(\lambda-1)}{1+\alpha+\alpha p_1(\lambda-1)} + \frac{d}{\Delta_p}\frac{\alpha(\lambda-1)(1+\alpha)}{(1+\alpha+\alpha p_1(\lambda-1))^2} \\
&\leq -\frac{d}{\Delta_p}\frac{\alpha(\lambda-1)}{(1+\alpha+\alpha p_1(\lambda-1))^2}(1+\alpha+\alpha p_1(\lambda-1)-(1+\alpha)) \\
&< 0
\end{aligned}$$

□

As the agent's preferences are more reference-dependent or more averse to losses, both (e_1, p_1) and (e_0, p_0) are personal equilibria for broader ranges of incentive intensity, and hence the region where multiple equilibria exist also enlarges. A more interesting exercise is to examine the effect of p_1, which we can interpret as a parameter representing the agent's ability.[18]

[18]More generally, let η be a parameter representing the agent's innate ability. The success probabilities p_0 and p_1 both depend on η, and assume the success probabilities are

Higher effort e_1 becomes more attractive to the agent because of higher success probability by choosing e_1. This first effect works to reduce both β_0 and β_1 while it affects β_0 more than β_1 because $\beta_0 > \beta_1$. For (e_1, p_1), there is an additional second effect of increasing p_1 via the reference point itself: Higher p_1 raises the probability of the reference being success (and hence the probability of loss), as well as reduces the probability of the reference being failure (and hence the probability of gain). Because of loss aversion, the former change dominates, which change in turn increases the marginal benefit from higher p_1. The additional effect thus also works to reduce β_1. However, this additional effect is not large enough to upset the smaller first effect on β_1 than on β_0. The region where there are multiple equilibria therefore becomes smaller as the agent has a higher ability.

4.2 The Optimal Contract

Now consider the principal's problem of solving the optimal contract. We suppose the benefit of success to the principal is so large that she wants to implement effort e_1 with least costs. Later in this subsection we take the benefit into consideration as well. The principal's problem is then to minimize the expected payment $b_f + p_1\Delta_b$ subject to the agent's participation constraint

$$U(e_1, p_1 \mid e_1, p_1) \geq \bar{u}, \tag{PC}$$

and a condition that the agent chooses (e_1, p_1) as a personal equilibrium, which we call the agent's incentive compatibility constraint (IC). (PC) implies that after observing the principal's offer, the agent anticipates his reference point and effort choice. We assume that the agent's reservation payoff

increasing in η: $p_i'(\eta) > 0$ for $i = 0, 1$. Furthermore, assume that $p_i(\eta)$ exhibits strictly increasing differences in (i, η): $p_1'(\eta) - p_0'(\eta) > 0$. In other words, ability and effort are complementary. We can then prove the results similar to those in Proposition 2: β_0, β_1, and $\beta_1 - \beta_0$ are decreasing in η.

is positive ($\bar{u} > 0$): If $\bar{u} = 0$, we need to take care of some boundary cases that complicate the analysis a bit. However, the results do not essentially change.

The principal faces a problem of multiple equilibria, but suppose for a moment that Δ_b can implement (e_1, p_1) as a personal equilibrium. For example, if the principal sets Δ_b a little above β_0, then she can guarantee the agent to play (e_1, p_1) since it is the only personal equilibrium.

Given such a Δ_b, the principal wants to minimize the expected payment. By (PC), the expected payment is bounded from below:

$$b_f + p_1\Delta_b \geq d + \bar{u} + p_1(1-p_1)\alpha(\lambda-1)\Delta_b. \tag{7}$$

(7) is satisfied at $b_f = 0$ when

$$\Delta_b \geq \frac{d+\bar{u}}{p_1}\frac{1}{1-(1-p_1)\alpha(\lambda-1)} \tag{8}$$

holds.[19] The optimal contract is then $(b_s, b_f) = (\Delta_b, 0)$ and the principal's expected payment is $p_1\Delta_b$. On the other hand, if (7) does not hold under $b_f = 0$, b_f is determined by the condition such that the inequality in (7) is replaced by equality. The principal's expected payment is then equal to the right-hand side of (7). Note that the principal's expected payment is increasing in Δ_b: If (7) does not bind, the agent earns rent, which is increasing in Δ_b; and if (7) binds, the principal's expected payment is equal to the right-hand side of (7), which is increasing in Δ_b.

We now study optimal contracts by explicitly analyzing the problem of multiple equilibria. Analysis depends on how the agent forms his expectation, facing multiple personal equilibria. We can think of three possible assumptions.

[19] Throughout this subsection we assume $\alpha \leq 1$ and $\lambda \approx 2$ so as to satisfy $1-(1-p_1)\alpha(\lambda-1) > 0$.

(i) The principal chooses the personal equilibrium she prefers, and the agent forms his expectation consistent to the chosen equilibrium.

(ii) The agent chooses the personal equilibrium he prefers.

(iii) The agent forms his expectation consistent to his inherent type: The *pessimistic* type chooses (e_0, p_0) while the *optimistic* type chooses (e_1, p_1).

If the first assumption is made, the optimal contract satisfies $\Delta_b = \beta_1$ since β_1 is the lowest incentive intensity such that the principal's preferred expectation (e_1, p_1) becomes an equilibrium. On the other hand, if we adopt the second assumption, the principal has virtually no role, except contract choice, in forming the agent's expectation, although the principal can still implement (e_1, p_1), possibly with Δ_b lower than β_0, under some conditions.[20]

In the rest of this subsection, we study the principal's optimal contract under the third assumption. We assume that the principal cannot distinguish between two types, and the proportion of the pessimistic (optimistic) agents is $q > 0$ (respectively $1 - q > 0$). We focus on contracts that induce at least the optimistic type to choose e_1. We further assume $p_0 = 0$ in order to simplify the analysis, so that $\Delta_p = p_1$.

We study three kinds of contractual arrangements that are different concerning how the principal deals with the pessimistic type: (a) *uniquely implementable contracts*; (b) *screening contracts*; and (c) *non-screening contracts*.

Uniquely implementable (UI) contracts induce not only the optimistic type but also the pessimistic type to choose e_1. The principal then must set Δ_b a little above β_0 so that (e_1, p_1) becomes the unique personal equilibrium.

[20] Later in section 5 we will provide such conditions. See Proposition 4.

By (8), the optimal UI contract is $(b_s, b_f) \approx (\beta_0, 0)$ if

$$\beta_0 > \frac{d+\bar{u}}{p_1}\frac{1}{1-(1-p_1)\alpha(\lambda-1)}$$

holds. Substituting $\beta_0 = (d/\Delta_b)(1+\alpha\lambda)/(1+\alpha+\alpha p_0(\lambda-1))$ and $p_0 = 0$, and solving for p_1 yield the following condition:

$$p_1 > \frac{D+\alpha\lambda}{1+\alpha\lambda}, \tag{9}$$

where $D \equiv (1+\alpha)\bar{u}/(\alpha(\lambda-1)d)$. If (9) hold, then the optimal UI contact is $(b_s, b_f) \approx (\beta_0, 0)$ and the principal's expected payment is approximately $p_1\beta_0$. If (9) does not hold, (PC) must bind at optimum, and the principal's expected payment is approximately equal to the right-hand side of (7) with $\Delta_b = \beta_0$. In summary, the expected payment under the optimal UI contract is approximately equal to c^{UI} defined as follows:

$$c^{UI} = \begin{cases} p_1\beta_0 & \text{if } p_1 > (D+\alpha\lambda)/(1+\alpha\lambda), \\ d+\bar{u}+p_1(1-p_1)\alpha(\lambda-1)\beta_0 & \text{otherwise.} \end{cases}$$

The second kind of contracts, *screening contracts*, induces only the optimistic agent to participate. For this aim, we have to impose the following additional condition, *screening condition*, to the principal's optimization problem:

$$\bar{u} > U(e_0, p_0 \mid e_0, p_0). \tag{SC}$$

A necessary condition for (PC) and (SC) to hold is that the optimistic agent earn a higher expected payoff than the pessimistic agent:

$$\begin{aligned} & U(e_1, p_1 \mid e_1, p_1) > U(e_0, p_0 \mid e_0, p_0) \\ \Rightarrow \quad & \Delta_b > \frac{d}{p_1}\frac{1}{1-(1-p_1)\alpha(\lambda-1)} \equiv \hat{\beta} \end{aligned} \tag{10}$$

This $\hat{\beta}$ is the lowest bound for Δ_b when the principal designs a contract screening out the pessimistic agent. Note that $\hat{\beta} > \beta_1$ always holds, while

$\hat{\beta} < \beta_0$ holds if and only if the following condition is satisfied:

$$p_1 > \frac{\alpha\lambda}{1+\alpha\lambda}. \tag{11}$$

Since $\bar{u} > 0$, (8) does not hold at $\Delta_b \approx \hat{\beta}$ and hence (PC) binds and condition (SC) holds at the optimal screening contract with $\Delta_b \approx \hat{\beta}$. The principal's expected payment is then approximately equal to

$$c^S = (1-q)\{d + \bar{u} + p_1(1-p_1)\alpha(\lambda-1)\hat{\beta}\}.$$

Under the third kind, *non-screening contracts*, the principal attempts to reduce the incentive intensity all the way down to β_1, and not only the optimistic agent but also the pessimistic agent participates. Since $\beta_1 < \hat{\beta}$, $U(e_1, p_1 \mid e_1, p_1) < U(e_0, p_0 \mid e_0, p_0)$ holds at $\Delta_b = \beta_1$. It is also easy to find that (8) does not hold at $\Delta_b = \beta_1$, and hence at the optimal non-screening contract with $\Delta_b = \beta_1$, the participation constraint (PC) binds for the optimistic agent and the principal's expected payment to the optimistic type is equal to $d + \bar{u} + p_1(1-p_1)\alpha(\lambda-1)\beta_1$.

However, the principal also has to pay for the pessimistic agent under the non-screening contract. Note that $U(e_0, p_0 \mid e_0, p_0) > \bar{u}$ since (PC) for the optimistic type is binding ($U(e_1, p_1 \mid e_1, p_1) = \bar{u}$) and $U(e_1, p_1 \mid e_1, p_1) < U(e_0, p_0 \mid e_0, p_0)$. This implies that the participation constraint for the pessimistic agent does not bind at optimum, and the principal's payment is equal to $b_f = d + \bar{u} + p_1(1-p_1)\alpha(\lambda-1)\beta_1 - p_1\beta_1$. The principal's expected payment is hence equal to

$$\begin{aligned} c^{NS} &= d + \bar{u} + p_1(1-p_1)\alpha(\lambda-1)\beta_1 - qp_1\beta_1 \\ &= d + \bar{u} + p_1[(1-p_1)\alpha(\lambda-1) - q]\beta_1 \end{aligned}$$

We now compare three kinds of contracts and derive the optimal contract. The following proposition compares the principal's expected payments among three contracts. The proof is provided in appendix.

Proposition 3. (a) If $p_1 > \alpha\lambda/(1+\alpha\lambda)$, then there exists $\tilde{q} \in (0,1)$ such that

(a1) $c^S < c^{NS} < c^{UI}$ if $q > \tilde{q}$;

(a2) $c^{NS} < c^S < c^{UI}$ if $q < \tilde{q}$.

(b) If $p_1 \leq \alpha\lambda/(1+\alpha\lambda)$, then there exists $\tilde{q}$ and $\bar{q}$ satisfying $0 \leq \bar{q} < \tilde{q} < 1$ such that

(b1) $c^S < c^{NS} < c^{UI}$ if $q > \tilde{q}$;

(b2) $c^{NS} < c^S < c^{UI}$ if $\bar{q} < q < \tilde{q}$;

(b3) $c^{NS} < c^{UI} < c^S$ if $q < \bar{q}$.

Proposition 3 implies that the lowest expected payment be attained either at the optimal screening contract or the optimal non-screening contract. The UI contract is costly because it induces both types of the agents to choose e_1 as a personal equilibrium, and hence must compensate for their loss aversion. It also must leave some rents to both types if $p_1 > (D+\alpha\lambda)/(1+\alpha\lambda)$.

The comparison between c^S and c^{NS} depends on the proportion of the pessimistic type q. First note that both c^S and c^{NS} are decreasing in q: c^S is decreasing in q because the agent is more likely to be screened out; and c^{NS} is decreasing in q since the expected payment to the pessimistic type is lower than that to the optimistic type. However, the pessimistic type earns rents under the optimal non-screening contract, while the optimistic type is paid more under the optimal screening contract because of the higher incentive intensity. The expected payment is hence lower under the screening contract than under the non-screening contract for q sufficiently high.

We have so far ignored the benefit of success. Now suppose the benefit is $B > 0$ if the project succeeds, while it is zero if the project fails. The comparison between the optimal screening contract and the optimal non-screening contract does not change since under either contract, the expected benefit to the principal is $(1-q)p_1B$. However, the optimal UI contract has an advantage since both types are induced to choose e_1 and hence the expected benefit is p_1B. The optimal contract depends on the magnitude of B. If it is sufficiently large, the UI contract is optimal. If B is small, then either the screening contract or the non-screening contract is optimal, following the conditions in Proposition 3.

5 The Pygmalion and Galatea Effects

Based on the analysis in the previous section, we formalize the Pygmalion and Galatea effects. Since the literature on this theme mainly focuses on effects of the subjects' abilities under given incentive schemes, we also restate our results in terms of a measure of the agent's ability, assuming that $\Delta_b > 0$ is given and the participation constraint is satisfied. We also continue to make a simplifying assumption $p_0 = 0$, and study how personal equilibria change with the agent's ability measured by p_1.

First, condition (PE1) for e_1 to become a personal equilibrium is rewritten as follows:

$$p_1[1+\alpha+\alpha(\lambda-1)p_1] \geq \frac{d}{\Delta_b}(1+\alpha) \qquad \text{(PE1a)}$$

Defining by π_1 the probability p_1 that satisfies (PE1a) with equality, we can show that π_1 has the following properties:

(i) $\pi_1 > 0$,

(ii) $\pi_1 < 1$ if $\Delta_b > (1+\alpha)d/(1+\alpha\lambda)$, and

(iii) If $p_1 \geq \pi_1$, then (e_1, p_1) is a personal equilibrium.

Similarly, condition (PE0) for e_0 to become a personal equilibrium is rewritten as follows:

$$p_1(1+\alpha) \leq \frac{d}{\Delta_b}(1+\alpha\lambda) \qquad \text{(PE0a)}$$

Denoting by π_0 the probability p_1 satisfying (PE0a) with equality, we can also show that π_0 has the following properties:

(i) $\pi_0 > 0$,

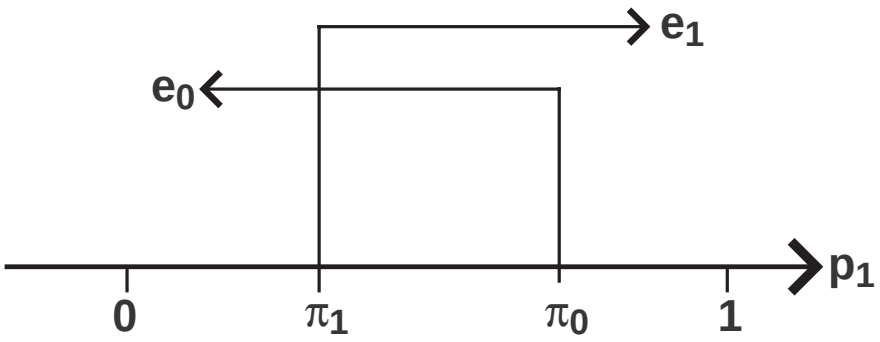
(ii) $\pi_0 < 1$ if $\Delta_b > (1+\alpha\lambda)d/(1+\alpha)$,

(iii) $\pi_0 \geq \pi_1$, in particular, $\pi_0 > \pi_1$ if $\Delta_b > (1+\alpha)d/(1+\alpha\lambda)$, and

(iv) If $p_1 \leq \pi_0$, then (e_0, p_0) is a personal equilibrium.

We can then identify three ranges of p_1 when $\Delta_b > (1+\alpha\lambda)d/(1+\alpha)$ (see Figure 2 where we denote personal equilibria only by effort level): (e_1, p_1) is a personal equilibrium in region $\pi_1 \leq p_1$ while (e_0, p_0) is a personal equilibrium in region $p_1 \leq \pi_0$. There exist multiple equilibria in region $\pi_1 \leq p_1 \leq \pi_0$. The formal results are summarized as the following corollary.

Figure 2: Personal Equilibria and the Agent's Ability



Corollary 1. When incentive intensity Δ_b is given, there are three ranges of p_1 that characterize personal equilibria.

(a) If $p_1 > \pi_0$, then (e_1, p_1) is the only personal equilibrium. This region exists if $\Delta_b > (1 + \alpha\lambda)d/(1 + \alpha)$.

(b) If $p_1 < \pi_1$, then (e_0, p_0) is the only personal equilibrium. This region always exists.

(c) If $\pi_1 \leq p_1 \leq \pi_0$, then both (e_0, p_0) and (e_1, p_0) are personal equilibria. This region exists if $\Delta_b > (1 + \alpha)d/(1 + \alpha\lambda)$.

Our formulation naturally explains the Galatea effect, one type of "self-fulfilling prophecy" which means that the agent's self-expectation about his performance determines his actual performance: If the agent thinks he can succeed with high probability, then he can actually succeed with the same high probability. More precisely, Corollary 1 states that the Galatea effect prevails if the agent's ability is sufficiently high ($p_1 \geq \pi_1$). On the other hand, if the agent's ability is low enough ($p_1 < \pi_1$), high expectation is not consistent to his actual choice, and hence no Galatea effect works.[21]

The Pygmalion effect, on the other hand, involves the principal. First, if the agent's ability is sufficiently high ($p_1 > \pi_0$), the principal's expectation does not play any role since the high effort is the only consistent reference expectation for the agent. Second, if the agent's expectation is low ($p_1 < \pi_1$), the principal's expectation does not play any role, either, since only the low effort is consistent to the agent's actual choice.

Finally, the principal's expectation could play a role for the agent with the intermediate levels of ability ($\pi_1 \leq p_1 \leq \pi_0$). There are multiple personal

[21] One could instead say the *Golem effect* exists in this case. The Golem effect is the Pygmalion effect in a negative direction: The Golem effect works when the low expectation induces low performance.

equilibria in this range, and hence it is important to specify how we assume the agent forms his expectation facing multiple equilibria. In subsection 4.2, we have considered three possible assumptions.

(i) The principal chooses the personal equilibrium she prefers, and the agent forms his expectation consistent to the chosen equilibrium.

(ii) The agent chooses the personal equilibrium he prefers.

(iii) The agent forms his expectation consistent to his inherent type: The *pessimistic* type chooses (e_0, p_0) while the *optimistic* type chooses (e_1, p_1).

If we adopt the first assumption, the principal's expectation works as follows. Suppose the agent initially has low expectation about his performance when low expectation is in equilibrium. The principal then informs the agent of her high expectation. The agent then examines whether such high expectation is consistent with his actual choice. If not so, he stays with the low expectation. However, if that expectation is consistent with his choice, he accepts the high expectation. Then, by the Galatea effect, the agent who has changed his expectation from low to high actually chooses high effort. This change can happen if the agent's ability is in the intermediate range $\pi_1 \leq p_1 \leq \pi_0$. The finding that the Pygmalion effect is important for the agent with the intermediate levels of ability is consistent with some evidence reported in Livingston (1969) mentioned in section 1.

The second assumption that the agent himself chooses his preferred personal equilibrium implies, of course, that there is no role for the principal's expectation, and hence the Pygmalion effect does not work. However, we could think of an intermediate case between the first and the second assumptions, where the principal attempts to make the agent attend to a particular

personal equilibrium under the condition that the agent may consciously choose his preferred equilibrium. The agent's payoff difference is calculated as follows.

$$\begin{aligned} &U(e_1, p_1 \mid e_1, p_1) - U(e_0, p_0 \mid e_0, p_0) \\ &\quad = p_1[1 - \alpha(\lambda - 1) + \alpha(\lambda - 1)p_1]\Delta_b - d \end{aligned} \tag{12}$$

Let us define $\hat{\pi}$ as the probability p_1 that satisfies (12) with equality. The agent prefers the personal equilibrium (e_1, p_1) to (e_0, p_0) if $p_1 \geq \hat{\pi}$. Moreover, if $\hat{\pi} < \pi_0$, then the Pygmalion effect may work for the agent whose ability is in the range between $\hat{\pi}$ and π_0 because such an agent prefers to change from e_0 to e_1 when the principal informs him of her high expectation. This range exists if incentive intensity is high enough to satisfy $\Delta_b > (1+\alpha)d/(1+\alpha\lambda)$ (so that $\pi_1 < \pi_0$ holds), but not too high. The results are provided formally in the following proposition.

Proposition 4.

(a) If $p_1 < \hat{\pi}$, then $U(e_0, p_0 \mid e_0, p_0) > U(e_1, p_1 \mid e_1, p_1)$.

(b) If $p_1 \geq \hat{\pi}$, then $U(e_0, p_0 \mid e_0, p_0) \leq U(e_1, p_1 \mid e_1, p_1)$. In addition, there exists a range of multiple equilibria with this property if $(1+\alpha)d/(1+\alpha\lambda) < \Delta_b < (1+\alpha\lambda)^2 d/(\alpha\lambda(1+\alpha))$.

Proof. Since most of the results are obvious from (12), we only show the proof of $\hat{\pi} < \pi_0$ if $\Delta_b < (1+\alpha\lambda)^2 d/(\alpha\lambda(1+\alpha))$. To show this, note that $U(e_1, p_1 \mid e_1, p_1) - U(e_0, p_0 \mid e_0, p_0) > 0$ holds if

$$p_1[1 - \alpha(\lambda - 1) + \alpha(\lambda - 1)p_1] > d/\Delta_b. \tag{13}$$

The left-hand side of (13) is a quadratic convex function of p_1 and is zero at $p_1 = 0$. We can then find the condition for Δ_b to satisfy $\hat{\pi} < \pi_0$ by

substituting $p_1 = \pi_0 = (d/\Delta_b)((1+\alpha\lambda)/(1+\alpha))$ into the left-hand side of (13), which leads to $\Delta_b < (1+\alpha\lambda)^2 d/(\alpha\lambda(1+\alpha))$. □

If the agent's ability is sufficiently small ($p_1 < \hat{\pi}$), then the agent always prefers (e_0, p_0) to (e_1, p_1) in the region with multiple equilibria, and hence the principal's attempt to make the agent attend to (e_1, p_1) may not be effective. On the other hand, if p_1 is high enough ($p_1 \geq \hat{\pi}$), (e_1, p_1) may be implementable through the principal's influence on the agent's expectation. We can hence say that the Pygmalion effect can work in this case. The reason why this effect does not work if the incentive intensity is too high is that such incentives do not satisfy (PE0a).

The third assumption is that the agent forms his expectation consistent to his inherent type. As in subsection 4.2, suppose there are two types (pessimistic type and optimistic type) of agents and consider screening and non-screening contracts. Then if the principal can change the agent's type from the pessimistic to the optimistic type thorough her expectation, we may interpret it as the Pygmalion effect. When the principal offers a screening contract, only the optimistic agent is willing to participate and choose e_1. Thus, if the principal is able to change the agent's expectation to be optimistic, she will benefit from it by making the agent's participation and implementation of e_1 more likely. On the other hand, when the principal offers a non-screening contract, the principal can further reduce incentive intensity to β_1, but not only the optimistic type but also the pessimistic type is willing to participate in such a contract, and only the former type will choose e_1. If, however, the principal can change the agent's type so as to make the optimistic type more likely, then her payoff increases since e_1 is more likely to be chosen for a given incentive intensity.

6 Concluding Remarks

In this paper we introduce RDP into an agency model to study interaction between reference dependence and incentives, and apply the results to both the Pygmalion and Galatea effects. We first show that the agent's higher expectation enables the principal to implement high effort with lower-powered incentives than when the agent does not exhibit RDP. Our agent evaluates his future choice based on his expectation as a reference point. The agent with higher self-expectation is thus going to perform better. We interpret this as the Galatea effect. We also show that when the agent's ability is intermediate, multiple equilibria exist. The principal wants to make the agent attend to high effort in the region with multiple equilibria, and we interpret the Pygmalion effect as an equilibrium selection device.

In this paper, we have explained the Pygmalion and Galatea effects under the assumption that both the principal and the agent know the agent's ability. One could alternatively study the Pygmalion effect under the condition that only the principal knows the true ability of the agent. The Pygmalion effect is then an effect of communication. The principal prefers to tell the agent that his ability is high, if the agent believes it. As mentioned in the introduction, the main issue of this alternative explanation is then how the principal can make her announcement credible, and we could analyze the problem in a fashion similar to Benabou and Tirole (2003) and Hermalin (1998).

Although we believe expectations play a crucial role in determining reference points, there are other plausible reference points. Status quo and endowments are obviously important determinants in many cases. Another interpretation of reference points is based on the goal setting theory in social psychology (Locke and Latham, 2002). The goal setting theory shows that

goals enhance performance by affecting reference points (see also Falk and Knell (2004) and Heath et al. (1999)).

It is easy to modify our model such that the reference point is the goal set by the principal. We can then show that the principal prefers to choose the highest effort level and success probability as the reference point, provided that the agent participates (Daido and Itoh, in progress). Choosing such a point maximizes the agent's feeling of loss when he shirks (choosing e_0), and hence enables the agent to choose e_1 by the lowest incentive Δ_b.[22] The principal might even suggest the agent be able to succeed with probability one, although such a point is inconsistent with the actual probability distribution.

This line of research may also be helpful to explain the Pygmalion effect. However, the reference point as the goal set by the principal does not always align with the agent's expectation which must be consistent with eventual outcomes. We believe that the agent plays a more active role in determining his reference point through his expectations.

As we have seen in the introduction, much research in social psychology, management, and so on, including laboratory and field studies, shows that the Pygmalion effect is significant. By contrast, there exists little economic research on the Pygmalion effect as far as we know. However, we believe that the Pygmalion effect also brings rich economic implications. We hope our paper stimulates future economic research, especially experimental one, on the Pygmalion effect, or more generally, self-fulfilling prophecy.

[22] On the other hand, Proposition 1 of Kőszegi and Rabin (2004) shows that the agent prefers the lowest reference point.

Appendix

Proof of Proposition 3.

We first compare c^S and c^{NS}. We have $c^S < c^{NS}$ if and only if

$$p_1 A(\hat{\beta} - \beta_1) < q(d + \bar{u} + p_1 A\hat{\beta} - p_1\beta_1)$$

where $A \equiv (1 - p_1)\alpha(\lambda - 1) < 1$. The terms in the parentheses in the right-hand side are positive since

$$d + \bar{u} + p_1 A\hat{\beta} - p_1\beta_1 > d + \bar{u} - p_1(1 - A)\beta_1 > 0.$$

The first inequality comes from $\hat{\beta} > \beta_1$, and the second inequality holds because (8) does not hold at $\Delta_b = \beta_1$, and hence

$$d + \bar{u} > p_1(1 - A)\beta_1. \tag{A1}$$

We thus obtain $c^S < c^{NS}$ if and only if

$$q > \tilde{q} \equiv \frac{p_1 A(\hat{\beta} - \beta_1)}{d + \bar{u} + p_1 A\hat{\beta} - p_1\beta_1} \tag{A2}$$

holds. The difference between the denominator and the numerator is calculated as

$$(d + \bar{u} + p_1 A\hat{\beta} - p_1\beta_1) - p_1 A(\hat{\beta} - \beta_1) = d + \bar{u} - p_1(1 - A)\beta_1 > 0$$

by (A1), and hence $0 < \tilde{q} < 1$ is satisfied.

We next compare c^{UI} with c^{NS}. First, suppose $p_1 > (D + \alpha\lambda)/(1 + \alpha\lambda)$ and hence

$$p_1(1 - A)\beta_0 > d + \bar{u}. \tag{A3}$$

Then by $\beta_0 > \beta_1$ and (A3),

$$\begin{aligned} c^{UI} - c^{NS} &= qp_1\beta_1 - (d + \bar{u} + p_1 A\beta_1 - p_1\beta_0) \\ &> qp_1\beta_1 - (d + \bar{u} - p_1(1 - A)\beta_0) > 0. \end{aligned}$$

Next suppose $p_1 \leq (D + \alpha\lambda)/(1 + \alpha\lambda)$. Then

$$c^{UI} - c^{NS} = qp_1\beta_1 + p_1A(\beta_0 - \beta_1) > 0.$$

The expected payment is thus always smaller under the optimal non-screening contract than under the optimal UI contract.

Finally, we compare c^S and c^{UI}. Suppose first $p_1 > (D + \alpha\lambda)/(1 + \alpha\lambda)$ Then $c^S < c^{UI}$ if and only if

$$q > \frac{d + \bar{u} + p_1A\hat{\beta} - p_1\beta_0}{d + \bar{u} + p_1A\hat{\beta}}$$

holds. However, the numerator of the right-hand side is negative by $\hat{\beta} < \beta_0$ and (A3). Hence $c^S < c^{UI}$ always holds.

Next suppose $p_1 \leq (D + \alpha\lambda)/(1 + \alpha\lambda)$. Then $c^S < c^{UI}$ if and only if

$$q > \bar{q} \equiv \frac{p_1A(\hat{\beta} - \beta_0)}{d + \bar{u} + p_1A\hat{\beta}}$$

holds. Note that $\bar{q} > 0$ if $\hat{\beta} > \beta_0$, and the sign of $\tilde{q} - \bar{q}$ is equal to the sign of

$$\begin{aligned} &p_1A(\hat{\beta} - \beta_1)(d + \bar{u} + p_1A\hat{\beta}) - p_1A(\hat{\beta} - \beta_0)(d + \bar{u} + p_1A\hat{\beta} - p_1\beta_1) \\ &= p_1A(\beta_0 - \beta_1)(d + \bar{u} + p_1A\hat{\beta}) + p_1^2A(\hat{\beta} - \beta_0)\beta_1 > 0, \end{aligned}$$

and hence $\bar{q} < \tilde{q} < 1$. Furthermore, $\bar{q} < 0$ if and only if $\hat{\beta} < \beta_0$ or

$$p_1 > \frac{\alpha\lambda}{1 + \alpha\lambda} \tag{11}$$

holds. Therefore if $p_1 > \alpha\lambda/(1 + \alpha\lambda)$, then $c^S < c^{UI}$ always holds, and if $p_1 \leq \alpha\lambda/(1 + \alpha\lambda)$, then $c^S < c^{UI}$ if and only if $q > \bar{q}$.

The analysis given above leads to the conclusion of Proposition 3. (Q.E.D.)

References

[1] Bénabou, R. and J.Tirole, 2003, Intrinsic and Extrinsic Motivation, *Review of Economic Studies* 70, 489-520.

[2] Daido, K. and H. Itoh, in progress, Goal Setting and Incentives.

[3] Falk, A. and M. Knell, 2004, Choosing the Joneses: Endogenous Goals and Reference Standards, *Scandinavian Journal of Economics* 106, 417-435.

[4] Goddard, R.W., 1985, The Pygmalion Effect, *Personnel Journal* 64, 10-16.

[5] Heath, C., R. P. Larrick and G. Wu, 1999, "Goals as Reference Points," *Cognitive Psychology* 38, 79–109.

[6] Hermalin, Benjamin, 1998, "Toward an Economic Theory of Leadership: Leading By Example," *American Economic Review* 88, 1188–1206.

[7] Jussim, L., 1986, Self-fulfilling prophecies: a theoretical and integrative review, *Psychological Review* 93, 429-445.

[8] Kahneman, D. and A. Tversky, 1979, Prospect Theory: An Analysis of Decision under Risk, *Econometrica* 47, 263-291.

[9] Kierein, N. M. and M.A. Gold, 2000, Pygmalion in work organizations: a meta-analysis, *Journal of Organizational Behaviour* 21, 913-928.

[10] Kőszegi, B. and M. Rabin, 2004, A Model of Reference-Dependent Preferences, mimeo.

[11] Livingston, J.S., 1969, Pygmalion in Management, *Harvard Business Review* 47, 81-89.

[12] Locke, E. A. and G. P. Latham, 2002, Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey, *American Psychologist* 57, 705–717.

[13] McNatt, D.B., 2000, Ancient Pygmalion Joins Contemporary Management: A Meta-Analysis of the Result, *Journal of Applied Psychology* 85, 314-322.

[14] Masatlioglu, Y. and E.A. Ok, 2005, Rational Choice with Status Quo Bias, *Journal of Economic Theory* 121, 1–29.

[15] Munro, A. and R. Sugden, 2003, On the theory of reference-dependent preferences, *Journal of Economic Behavior and Organization* 50, 407-428.

[16] Rosenthal, R. and L. Jacobson, 1968, *Pygmalion in the Classroom: Teacher Expectations and Pupils' Intellectual Development*. Holt, Reinhart and Winston: New York.

[17] Sagi, J.S., 2004, Anchored Preference Relation, mimeo.

[18] Sugden, R., 2003, Reference-dependent subjective expected utility, *Journal of Economic Theory* 111, 172-191.

[19] Tversky, A. and D. Kahneman, 1991, Loss Aversion in Riskless Choice: A Preference-Dependent Model, *Quarterly Journal of Economics* 106, 1039–1061.