Daido, Kohei; Itoh, Hideshi

Working Paper

The Pygmalion effect : an agency model with reference dependent preferences

CESifo Working Paper, No. 1444

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Daido, Kohei; Itoh, Hideshi (2005) : The Pygmalion effect : an agency model with reference dependent preferences, CESifo Working Paper, No. 1444, Center for Economic Studies and Ifo Institute (CESifo), Munich

This Version is available at:
http://hdl.handle.net/10419/18808

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
THE PYGMALION EFFECT: AN AGENCY MODEL WITH REFERENCE DEPENDENT PREFERENCES

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 Köszegi 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 show that the principal’s expectation has an important role as an equilibrium selection device.

JEL Code: B49, D82, M12, M52, M54.

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

Kohei Daido  
Faculty of Economics  
Osaka Sangyo University  
3-1-1 Nakagaito  
Daito-shi, Osaka 574-8530  
Japan  
daido@eco.osaka-sandai.ac.jp

Hideshi Itoh  
Graduate School of Commerce and Management  
Hitotsubashi University  
2-1 Naka  
Kunitachi, Tokyo 186-8601  
Japan  
h.itoh@srv.cc.hit-u.ac.jp

Preliminary and incomplete. We are grateful to Kenichi Amaya and Munetomo Ando, and seminar participants at Osaka University Thursday Seminar 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, is gratefully acknowledged.
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 not in educational setting but in managerial setting.\(^2\) He argues that (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,” and (d) “subordinates, more often than not, appear to do what they believe they are expected to do.” Since Livingston (1969), many researchers have been

\(^1\)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).
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. A manager’s high expectation influences her attitude toward her subordinates, and such attitude has positive effects on subordinates’ self-expectancy. The subordinates’ enhanced self-expectancy then improves their performance. In this process, the part 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 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 PDP 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

---

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.

4 The seminal paper by Kahneman and Tversky (1979) explores issues on RDP. Recently, various models of RDP have been developed. For example, Masatilloglu and Ok (2003), Sagi (2004), and Sugden (2003) give axiomatic foundations for models of RDP.
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 (2003) and Sagi (2004) 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 two important features. First, a person’s reference point is her recent expectation, represented by a probability measure, over outcomes. Second, the model by Kőszegi and Rabin (2004) has a prominent feature that the reference point is endogenously determined by the person’s rational expectation. To this end, they define the personal equilibrium which requires that a person maximize his payoff given his rational expectations about outcomes, and hence the expectations themselves depend on his own anticipated behavior.

Following Kőszegi and Rabin (2004), 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

---

5We give a brief sketch of their model in the next section.
6In this sense, their model is similar to that of Sugden (2003). In Kőszegi and Rabin (2004), however, each outcome is compared to all outcomes in the support of the reference lottery, while in Sugden (2003), an outcome is compared only to the outcome that would have resulted from the reference lottery in the same state.
7Munro and Sugden (2003) and Falk and Knell (2004) also study the endogenous determination of reference points.
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 interpret this result as the Galatea effect. 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. In this respect, we can interpret the Pygmalion effect as the principal’s attempt to make the agent choose a particular personal equilibrium.

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. The second, more interesting approach is to focus on the role of information transmission by the principal. When the agent does not know his own productivity, the principal with high expectation may effectively transmit her private information on the agent’s productivity.\footnote{Note that the principal’s mere expression of her expectation is not credible to the agent, since it is just cheap talk.} Bénabou and Tirole (2003) is an example
along this line. 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, this kind of explanation disregards the crux of the Pygmalion effect. Firstly, 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. Then, the performance of the top group which consists of his best agents improved dramatically. However, the performance of the middle group which consists of his average agents also significantly improved. Surprisingly, this group increased its productivity by a higher percentage than the top group did. Secondly, 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 introduce the model of RDP based on Kőszegi and Rabin (2004). In section 3 we build a simple agency model with RDP. We analyze the personal equilibrium and study the optimal incentive scheme in the same section. In section 4 we extend our model to some directions. Section 5 concludes.

2 Reference Dependent Preferences

Following Kőszegi and Rabin (2004), we formulate the reference dependent nature of preferences as the following way. Let \( c = (c_1, \ldots, c_n) \) be a con-
sumption bundle of an agent, and \( r = (r_1, \ldots, r_n) \) be a reference consumption bundle. We define the agent’s overall payoff \( u(c \mid r) \) by

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

(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 gains and losses 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.

This 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

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

We further assume that each dimension is evaluated by the same function \( \mu(\cdot) \) of the difference of consumption from the reference level, evaluated by material payoff:

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

(2)

The “universal” gain-loss function \( \mu(\cdot) \) is assumed to have the following properties. They captures 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 \( \lim_{y \to 0} \mu'(y)/\lim_{y \to 0} \mu'(y) = \gamma > 1 \).
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 the current version of the paper, we isolate the effect of loss aversion by assuming $\mu(\cdot)$ is linear, and define

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

where $\alpha$, a positive constant, is the weight on the gain-loss payoff, and $\gamma > 1$ is the “coefficient of loss aversion” which is the same as $\gamma$ defined in A3.

3 A Simple Agency Model

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 $e_i$.

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 $u(w, e_i | \overline{w}, e_j)$ by

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

In Section 4, we consider a 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.$^{10}$

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 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 argue that most existing empirical work $^{10}$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.
on loss aversion, in which the status quo is exogenously given, can also be interpreted in terms of expectations as reference points. 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 will 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 \(p_j\). Then \((e_j, p_j)\) is a personal equilibrium if for \(i \neq j\),

\[
U(e_j, p_j | e_j, p_j) \geq U(e_i, p_i | e_j, p_j) \tag{5}
\]

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

\[
U(e_i, p_i | e_j, p_j) = b_f + p_i \Delta_b - d_i + p_i(1 - p_j)\mu(\Delta_b) + (1 - p_i)p_j\mu(-\Delta_b) + \mu(d_j - d_i). \tag{6}
\]

The first line of (6) is the expected material payoff. The second and third lines represent the gain-loss payoff under reference point \((e_j, p_j)\).\(11\) For example, suppose that the reference point is \((e_1, p_1)\), and the agent’s choice is

\(11\)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.
$(e_0, p_0)$. The agent enjoys gain $\alpha d$ (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 $\alpha \gamma \Delta b$ with probability $p_1$ since he was expecting to succeed with this probability. These gain and loss correspond to the second line.

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.

### 3.1 Analysis: 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 $(e_0, p_0)$.

First consider $(e_1, p_1)$. The relevant expected payoffs are calculated as follows:

$$U(e_1, p_1 \mid e_1, p_1) = b_f + p_1 \Delta b - d + p_1(1 - p_1)\alpha \Delta b - (1 - p_1)p_1 \alpha \gamma \Delta b$$

$$U(e_0, p_0 \mid e_1, p_1) = b_f + p_0 \Delta b + \alpha d + p_0(1 - p_1)\alpha \Delta b - (1 - p_0)p_1 \alpha \gamma \Delta b$$

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

$$\Delta b \geq \beta_1 = \frac{d}{\Delta b} \frac{1 + \alpha}{1 + \alpha p_1(\gamma - 1)} \quad \text{(PE1)}$$

Mixed strategies will be analyzed later in Section 4.
Next consider \((e_0, p_0)\). The following expected payoffs are relevant.

\[
U(e_0, p_0 | e_0, p_0) = b_f + p_0 \Delta_b + p_0(1 - p_0)\alpha \Delta_b - (1 - p_0)p_0\alpha \gamma \Delta_b
\]

\[
U(e_1, p_1 | e_0, p_0) = b_f + p_1 \Delta_b - d - \alpha \gamma d + p_1(1 - p_0)\alpha \Delta_b - (1 - p_1)p_0\alpha \gamma \Delta_b
\]

Then \((e_0, p_0)\) is a personal equilibrium if

\[U(e_0, p_0 | e_0, p_0) \geq U(e_1, p_1 | e_0, p_0),\]

or

\[
\Delta_b \leq \beta_0 \equiv \frac{d}{\Delta_p} \frac{1 + \alpha \gamma}{1 + \alpha + \alpha p_0 (\gamma - 1)} \quad \text{(PE0)}
\]

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” \(\Delta_b\) 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 \(\Delta_b\) lower than \(d/\Delta_p\), the critical value under the standard case. And the agent, expecting to choose \(e_0\), actually chooses \(e_0\) for \(\Delta_b\) higher than \(d/\Delta_p\).

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 $ad$. The positive effects dominate because of $\gamma > 1$, 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 $\alpha\gamma d$. 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 $\Delta_b$ in Proposition 1 (ii). An impor-
tant finding is that in the intermediate range, there are multiple personal
equilibria. In this range, the agent chooses $e_1$ if he expects to do so, while
he chooses $e_0$ if $e_0$ is his expected effort.

Figure 1: Personal Equilibria
The next proposition reports some comparative statics results.

**Proposition 2.**

(a) $\beta_0$ is increasing in $\alpha$ and $\gamma$, and decreasing in $p_1$.

(b) $\beta_1$ is decreasing in $\alpha$, $\gamma$ and $p_1$.

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

**Proof.** Results (a) and (b) directly follow from the definitions in (PE0) and (PE1). For (c), the comparative statics results for $\alpha$ and $\gamma$ are immediate from (a) and (b). By differentiating $\beta_0$ and $\beta_1$ by $p_1$, we obtain

\[
\frac{\partial \beta_0}{\partial p_1} - \frac{\partial \beta_1}{\partial p_1} = -\frac{d}{\Delta_p^2} \left( \frac{1 + \alpha \gamma}{1 + \alpha + \alpha p_0(\gamma - 1)} - \frac{1 + \alpha}{1 + \alpha + \alpha p_1(\gamma - 1)} \right)
\]

\[
+ \frac{d}{\Delta_p} \frac{\alpha(\gamma - 1)(1 + \alpha)}{(1 + \alpha + \alpha p_1(\gamma - 1))^2}
\]

\[
\leq -\frac{d}{\Delta_p} \frac{\alpha(\gamma - 1)}{1 + \alpha + \alpha p_1(\gamma - 1)} + \frac{d}{\Delta_p} \frac{\alpha(\gamma - 1)(1 + \alpha)}{(1 + \alpha + \alpha p_1(\gamma - 1))^2}
\]

\[
\leq -\frac{d}{\Delta_p} \frac{\alpha(\gamma - 1)}{(1 + \alpha + \alpha p_1(\gamma - 1))^2} (1 + \alpha + \alpha p_1(\gamma - 1) - (1 + \alpha))
\]

\[
< 0
\]

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. Higher effort $e_1$ becomes more attractive to the agent because of higher success probability by choosing $e_1$. This effect works to reduce both $\beta_0$ and $\beta_1$ while it affects $\beta_0$ more than $\beta_1$ because $\beta_0 > \beta_1$. For $(e_1, p_1)$, there is an additional 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 \( \beta_1 \). However, this additional
effect is not large enough to upset the smaller first effect on \( \beta_1 \) than on \( \beta_0 \).
The region where there are multiple equilibria therefore becomes smaller as
the agent has a higher ability.

### 3.2 Analysis: The Optimal Contract

Now consider the principal’s problem of solving the optimal contract. We
assume that the benefit of success to the principal is so large that she wants
to implement effort \( e_1 \) with least costs. The principal’s problem is then to
minimize the expected payment \( b_f + p_1 \Delta_b \) subject to the agent’s participation
constraint

\[
U(e_1, p_1 | e_1, p_1) \geq 0,
\]

(PC)

and a condition that the agent chooses \((e_1, p_1)\) as a personal equilibrium,
which we call IC (incentive compatibility constraint). (PC) implies that after
observing the principal’s offer, the agent anticipates his reference point and
effort choice. Although the reservation payoff is normalized to zero, all our
results continue to hold under a general reservation payoff \( \bar{U} \) with minor
modification.

Constraint (IC) faces a problem of multiple equilibria, but suppose for a
while that \( \Delta_b \) can implement \((e_1, p_1)\) as a personal equilibrium. For example,
if the principal sets \( \Delta_b \) a little above \( \beta_0 \), then she can guarantee the agent
to play \((e_1, p_1)\) since it is the only personal equilibrium.

Given \( \Delta_b \), the principal wants to minimize the expected payment. By
(PC), the expected payment is bounded from below since

\[ b_f + p_1 \Delta b \geq d + p_1(1 - p_1)\alpha \Delta b(\gamma - 1). \]  

(7)

If \( b_f = 0 \) satisfies (7), then \((b_s, b_f) = (\Delta b, 0)\) is the optimal contract. If (7) does not hold under \( b_f = 0 \), then \( b_f \) is determined by the condition such that the inequality in (7) is replaced by equality.

Note that the principal’s expected payment is increasing in \( \Delta b \): If (7) does not bind, the agent earns rent, which is increasing in \( \Delta b \); and if (7) binds, the principal’s expected payment is equal to the right-hand side of (7), which is increasing in \( \Delta b \).

3.3 The Pygmalion and Galatea Effects

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, then he is more likely to succeed. More precisely, Proposition 1 states that the Galatea effect prevails when \( \Delta b \geq \beta_1 \) holds.

The Pygmalion effect, on the other hand, involves the principal. The principal in our model can influence the personal equilibrium by changing incentive intensity \( \Delta b \). If she sets \( \Delta b \) a little above \( \beta_0 \), then the only personal equilibrium is \((e_1, p_1)\), and hence the agent expects to choose \( e_1 \) and succeed with probability \( p_1 \) (and he does). The principal could also play an important role in influencing the agent’s formation of the reference point. The principal prefers to reduce \( \Delta b \) further up to \( \beta_1 \), if she can make sure the agent plays \((e_1, p_1)\) in the range with multiple equilibria. We can thus say that the Pygmalion effect, meaning that the agent performs in accordance with the principal’s expectations, realizes in our model if the principal can induce the agent to attend to the “right” expectation.
Note that the principal’s expectation has this role in the region where multiple equilibria exit. By Proposition 2, this range increases with \( \alpha, \gamma \), and decreases with \( p_1 \), and hence the principal’s expectation is more likely to matter as the agent is more dependent on the reference point, more averse to losses, or the agent’s ability is lower. This last property is consistent with the empirical result that the Pygmalion effect works best for people whose initial performance level is low (Kierein and Gold, 2000; McNatt, 2000).

One potential obstacle against the principal’s attempt to make the agent attend to a particular reference point is that the agent himself may consciously choose his preferred equilibrium. The agent’s payoff difference is calculated as follows.

\[
U(e_1, p_1 \mid e_1, p_1) - U(e_0, p_0 \mid e_0, p_0)
= \Delta_b \Delta_p [1 + \alpha(\gamma - 1)(p_0 + p_1) - \alpha(\gamma - 1)] - d
\]

(8)

We thus obtain the following results.

**Proposition 3.**

(a) If \( p_1 \leq 1 - p_0 - (\alpha(\gamma - 1))^{-1} \), then \( U(e_1, p_1 \mid e_1, p_1) < U(e_0, p_0 \mid e_0, p_0) \).

(b) If \( p_1 > 1 - p_0 - (\alpha(\gamma - 1))^{-1} \), then \( U(e_1, p_1 \mid e_1, p_1) \geq U(e_0, p_0 \mid e_0, p_0) \) if and only if

\[
\Delta_b \geq \hat{\beta} \equiv \frac{d}{\Delta_b} \frac{1}{1 + \alpha(\gamma - 1)(p_0 + p_1) - \alpha(\gamma - 1)}.
\]

(9)

Threshold value \( \hat{\beta} \) is decreasing in \( p_1 \), and \( \hat{\beta} < \beta_0 \) for sufficiently large \( p_1 \).

**Proof.** (a) and (9) follow from the right-hand side of (8). (8) also shows that \( \hat{\beta} \) is decreasing in \( p_1 \). And \( \hat{\beta} < \beta_0 \) holds if and only if

\[
p_1 > \frac{\alpha \gamma}{1 + \alpha \gamma} (1 - p_0),
\]

(10)
which is true if $p_1 \to 1$.\footnote{\(\beta_1 < \hat{\beta}\) holds if and only if
\[
\frac{1 + \alpha}{\alpha}(1 - p_0) > p_1.
\]
This may or may not hold as $p_1 \to 1$.}

If $p_1$ is sufficiently small ($p_1 \leq 1 - p_0 - (\alpha(\gamma - 1))^{-1}$), or in the intermediate range ($p_1 > 1 - p_0 - (\alpha(\gamma - 1))^{-1}$ holds but (10) fails to hold), 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, and she may not be able to lower $\Delta_b$ below $\beta_0$.

If $p_1$ is large enough (so that (10) is true), \((e_1, p_1)\) may be implementable with incentive intensity lower than $\beta_0$, but implementation may be difficult for $\Delta_b$ below $\hat{\beta}$ if the agent chooses his preferred equilibrium.

\section{Extensions}

\subsection{Mixed Strategies}

In the analysis in the previous section, we have focused on pure strategies ($e_0$ or $e_1$). Now suppose the agent expects to choose $e_1$ with probability $q$ (and $e_0$ with $1 - q$). The agent’s expected payoff, when actually choosing $e_1$ with probability $s$, is given as follows.

\[
U(s \mid q) = b_f + p(s)\Delta_b - sd - s(1 - q)\alpha\gamma d + (1 - s)q\alpha d + p(s)(1 - p(q))\alpha\Delta_b - (1 - p(s))p(q)\alpha\gamma\Delta_b
\]

where $p(x)$ is the probability of success under mixed strategy $x$, and is given by $p(x) = xp_1 + (1 - x)p_0$. Given $\Delta_b$, \((q, p(q))\) is a personal equilibrium if $U(q \mid q) \geq U(s \mid q)$ for all $s$. This holds if and only if the agent is indifferent
between \(e_0\) and \(e_1\), or \(q\) satisfies

\[
\Delta_b = \frac{d}{\Delta_P} \frac{1 + \alpha \gamma - \alpha (\gamma - 1)q}{1 + \alpha + \alpha (\gamma - 1)p(q)}.
\] (11)

The right-hand side is decreasing in \(q\), and is equal to \(\beta_0\) if \(q = 0\), and is equal to \(\beta_1\) if \(q = 1\). A unique mixed strategy personal equilibrium thus exists in the region with multiple equilibria, that is, if \(\Delta_b\) satisfies \(\beta_1 < \Delta_b < \beta_0\).

Interestingly, as the principal provides stronger incentive intensity, in the mixed strategy equilibrium the agent is less likely to choose \(e_1\): Increasing \(\Delta_b\) leads to \(e_1\) being more attractive, and hence \(q\) must be reduced to make \(e_1\) and \(e_0\) indifferent.

Note that the agent’s expected payoff under the mixed strategy equilibrium is between the expected payoffs at two pure strategy equilibria. This is the reason we have focused on pure strategies in the main analysis: it is unlikely that either the principal or the agent is interested in playing the mixed strategy equilibrium.

### 4.2 Continuous Effort

Our main analysis is restricted to the simple case in which there are two outcomes and two effort levels. In this subsection we continue to assume the outcome of the project is either success or failure, while we relax the former restriction and instead assume that the set of feasible effort is continuous.

We suppose that the agent chooses effort \(e\) from \(E = [0, \bar{e}]\). For simplicity, the agent’s cost of effort is \(e\). The project succeeds with probability \(p(e)\), which we assume twice-continuously differentiable and satisfies \(p'(e) > 0\) and \(p''(e) < 0\) for all \(e\). When the agent expects to choose \(e\) (and succeed with probability \(p(e)\)) and actually chooses \(\hat{e} \leq e\), his expected payoff is
given as follows.

\[
U(\hat{e} \mid e) = bf + p(\hat{e})\Delta_b - \hat{e} \\
+ p(\hat{e})(1 - p(e))\alpha\Delta_b - (1 - p(\hat{e}))p(e)\alpha\gamma\Delta_b
\]
\[
+ \alpha(e - \hat{e}).
\]

The first-order condition for the agent to choose \( e \) is given by

\[
p'(e)(1 + \alpha + \alpha(\gamma - 1)p(e))\Delta_b \geq 1 + \alpha. \tag{12}
\]

Similarly, the agent’s expected payoff, when he chooses \( \hat{e} \geq e \), is

\[
U(\hat{e} \mid e) = bf + p(\hat{e})\Delta_b - \hat{e} \\
+ p(\hat{e})(1 - p(e))\alpha\Delta_b - (1 - p(\hat{e}))p(e)\alpha\gamma\Delta_b
\]
\[
- \alpha\gamma(\hat{e} - e).
\]

The first-order condition is given by

\[
p'(e)(1 + \alpha + \alpha(\gamma - 1)p(e))\Delta_b \leq 1 + \alpha\gamma. \tag{13}
\]

If the principal wants to implement \( e \), then by setting \( \Delta_b \) satisfying both (12) and (13), she can induce \( e \) to be a personal equilibrium for the agent. Although there is the multiple equilibrium problem as before, we here suppose that the agent attends to the equilibrium the principal prefers most, in order to explore the interaction between personal equilibria and incentives. Since the principal prefers lower \( \Delta_b \), it is optimal for her to set \( \Delta_b \) such that (12) holds with equality, and hence the following incentive intensity is optimal.

\[
\Delta_b = \frac{1}{p'(e)} \frac{1 + \alpha}{1 + \alpha + \alpha(\gamma - 1)p(e)} \tag{14}
\]

Differentiating \( \Delta_b \) with regard to \( e \) yields the following result.

\[
\text{sign} \left[ \frac{\partial \Delta_b}{\partial e} \right] = -\text{sign} \left[ \alpha(\gamma - 1)(p(e)p''(e) + (p'(e))^2) + p''(e)(1 + \alpha) \right] \tag{15}
\]
First consider the agent with the standard reference-independent preferences \((\alpha = 0)\). Then from (15) it is easy to see \(\Delta_b\) is increasing in \(e\): A more powerful incentive intensity is needed for implementation of a higher effort. This is a standard result. Now suppose the agent has RDP \((\alpha > 0)\). (15) then does not tell us whether or not \(\Delta_b\) is increasing. In other words, there is a possibility that the optimal incentive intensity decreases with effort.\(^{14}\) The agent expecting a higher effort as a reference point faces a higher chance of loss and a lower chance of gain. Because of loss aversion, the former effect dominates, and hence the agent is more motivated to choose a higher effort. If this effect is sufficiently large, the principal can save monetary incentives, and lower \(\Delta_b\) is optimal.

5 Concluding Remarks

In this paper we present the agency model with RDP to explain 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. 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 power of incentives 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.

It is crucial for us to take the viewpoint that the reference point is the agent’s expectation. However, there is an alternative interpretation of reference points based on the goal setting theory in social psychology.

\(^{14}\)We can construct an example such that \(\Delta_b\) is decreasing in all \(e\).
(Locke and Latham, 2002). The goal setting theory shows that goals enhance performance by affecting reference points. It is easy to modify our model such that the reference point is the goal set by the principal. We can then show that it is optimal for the principal to choose the highest effort level as the reference point. This line of research may also be helpful to explain the Pygmalion effect.

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.

References


1378 Theodore C. Bergstrom and John L. Hartman, Demographics and the Political Sustainability of Pay-as-you-go Social Security, January 2005

1379 Bruno S. Frey and Margit Osterloh, Yes, Managers Should Be Paid Like Bureaucrats, January 2005

1380 Oliver Hüls eig, Eric Mayer and Timo Wollmershäuser, Bank Loan Supply and Monetary Policy Transmission in Germany: An Assessment Based on Matching Impulse Responses, January 2005

1381 Alessandro Balestrino and Umberto Galmarini, On the Redistributive Properties of Presumptive Taxation, January 2005

1382 Christian Gollier, Optimal Illusions and Decisions under Risk, January 2005

1383 Daniel Mejía and Marc St-Pierre, Unequal Opportunities and Human Capital Formation, January 2005


1385 Ruslan Lukach, Peter M. Kort and Joseph Plasmans, Optimal R&D Investment Strategies with Quantity Competition under the Threat of Superior Entry, January 2005

1386 Alfred Greiner, Uwe Koeller and Willi Semmler, Testing Sustainability of German Fiscal Policy. Evidence for the Period 1960 – 2003, January 2005


1388 Emanuele Bacchiocchi and Alessandro Missale, Managing Debt Stability, January 2005

1389 Assar Lindbeck and Dirk Niepelt, Improving the SGP: Taxes and Delegation rather than Fines, January 2005

1390 James J. Heckman and Dimitriy V. Masterov, Skill Policies for Scotland, January 2005

1391 Emma Galli & Fabio Padovano, Sustainability and Determinants of Italian Public Deficits before and after Maastricht, January 2005

1392 Angel de la Fuente and Juan Francisco Jimeno, The Private and Fiscal Returns to Schooling and the Effect of Public Policies on Private Incentives to Invest in Education: A General Framework and Some Results for the EU, January 2005
Juan C. Conesa and Carlos Garriga, Optimal Response to a Demographic Shock, January 2005

Christian Gollier, Optimal Portfolio Management for Individual Pension Plans, February 2005

Ruslan Lukach, Joseph Plasmans and Peter M. Kort, Innovation Strategies in a Competitive Dynamic Setting, February 2005

Gebhard Kirchgässner, (Why) Are Economists Different?, February 2005


Gabrielle Demange, Free Choice of Unfunded Systems: A First Assessment, February 2005

Carlos Fonseca Marinheiro, Sustainability of Portuguese Fiscal Policy in Historical Perspective, February 2005


Frank Asche, Petter Osmundsen and Maria Sandsmark, Is It All Oil?, February 2005

Giacomo Corneo, Media Capture in a Democracy: The Role of Wealth Concentration, February 2005

A. Lans Bovenberg and Thijs Knaap, Ageing, Funded Pensions and the Dutch Economy, February 2005

Thiess Büttner, The Incentive Effect of Fiscal Equalization Transfers on Tax Policy, February 2005

Luisa Fuster, Ayşe İmrohoroğlu and Selahattin İmrohoroğlu, Personal Security Accounts and Mandatory Annuityzation in a Dynastic Framework, February 2005

Peter Claeys, Policy Mix and Debt Sustainability: Evidence from Fiscal Policy Rules, February 2005

James M. Malcomson, Supplier Discretion over Provision: Theory and an Application to Medical Care, February 2005

Thorvaldur Gylfason, Interview with Assar Lindbeck, February 2005

Christian Gollier, Some Aspects of the Economics of Catastrophe Risk Insurance, February 2005

Gebhard Kirchgässner, The Weak Rationality Principle in Economics, February 2005
1411 Carlos José Fonseca Marinheiro, Has the Stability and Growth Pact Stabilised? Evidence from a Panel of 12 European Countries and Some Implications for the Reform of the Pact, February 2005

1412 Petter Osmundsen, Frank Asche, Bård Misund and Klaus Mohn, Valuation of International Oil Companies – The RoACE Era, February 2005

1413 Gil S. Epstein and Shmuel Nitzan, Lobbying and Compromise, February 2005

1414 Marcel F. M. Canoy, Jan C. van Ours and Frederick van der Ploeg, The Economics of Books, February 2005


1416 George Kapetanios and M. Hashem Pesaran, Alternative Approaches to Estimation and Inference in Large Multifactor Panels: Small Sample Results with an Application to Modelling of Asset Returns, February 2005

1417 Samuel Mühlemann, Jürg Schweri, Rainer Winkelmann and Stefan C. Wolter, A Structural Model of Demand for Apprentices. February 2005

1418 Giorgio Brunello and Lorenzo Rocco, Educational Standards in Private and Public Schools, February 2005

1419 Alex Bryson, Lorenzo Cappellari and Claudio Lucifora, Why so Unhappy? The Effects of Unionisation on Job Satisfaction, March 2005

1420 Annalisa Luporini, Relative Performance Evaluation in a Multi-Plant Firm, March 2005

1421 Giorgio Bellettini and Carlotta Berti Ceroni, When the Union Hurts the Workers: A Positive Analysis of Immigration Policy, March 2005

1422 Pieter Gautier, Michael Svarer and Coen Teulings, Marriage and the City, March 2005

1423 Ingrid Ott and Stephen J. Turnovsky, Excludable and Non-Excludable Public Inputs: Consequences for Economic Growth, March 2005

1424 Frederick van der Ploeg, Back to Keynes?, March 2005


1426 Hans Pitlik, Friedrich Schneider and Harald Strotmann, Legislative Malapportionment and the Politicization of Germany’s Intergovernmental Transfer System, March 2005

1427 Konstantinos Angelopoulos and Apostolis Philippopoulos, The Role of Government in Anti-Social Redistributive Activities, March 2005
1428 Ansgar Belke and Daniel Gros, Asymmetries in the Trans-Atlantic Monetary Policy Relationship: Does the ECB follow the Fed?, March 2005

1429 Sören Blomquist and Luca Micheletto, Optimal Redistributive Taxation when Government’s and Agents’ Preferences Differ, March 2005

1430 Olof Åslund and Peter Fredriksson, Ethnic Enclaves and Welfare Cultures – Quasi-Experimental Evidence, March 2005


1432 Peter Egger, Stefan Gruber, Mario Larch and Michael Pfaffermayr, Knowledge-Capital Meets New Economic Geography, March 2005

1433 George Economides and Apostolis Philippopoulos, Should Green Governments Give Priority to Environmental Policies over Growth-Enhancing Policies?, March 2005

1434 George W. Evans and Seppo Honkapohja, An Interview with Thomas J. Sargent, March 2005

1435 Helge Berger and Volker Nitsch, Zooming Out: The Trade Effect of the Euro in Historical Perspective, March 2005

1436 Marc-Andreas Muendler, Rational Information Choice in Financial Market Equilibrium, March 2005

1437 Martin Kolmar and Volker Meier, Intra-Generational Externalities and Inter-Generational Transfers, March 2005

1438 M. Hashem Pesaran and Takashi Yamagata, Testing Slope Homogeneity in Large Panels, March 2005

1439 Gjermund Nese and Odd Rune Straume, Industry Concentration and Strategic Trade Policy in Successive Oligopoly, April 2005

1440 Tomer Blumkin and Efraim Sadka, A Case for Taxing Education, April 2005

1441 John Whalley, Globalization and Values, April 2005

1442 Denise L. Mauzerall, Babar Sultan, Namsoung Kim and David F. Bradford, Charging NOx Emitters for Health Damages: An Exploratory Analysis, April 2005

1443 Britta Hamburg, Mathias Hoffmann and Joachim Keller, Consumption, Wealth and Business Cycles in Germany, April 2005

1444 Kohei Daido and Hideshi Itoh, The Pygmalion Effect: An Agency Model with Reference Dependent Preferences, April 2005