A modified inspector leadership game with psychological factors

Masakazu Fukuzumi*

February 8, 2007

Abstract

We formulate a monitoring model which is a modified inspector leadership game where a principal (an inspector) monitors an effort level chosen by an agent (an inspectee). We introduce psychological factors (a sense of guilty and an impulse to deceive) into the modified inspector leadership game and examine impacts of these psychological factors on an error probability that the principal conducts a *costly* investigation into an effort level chosen by the agent *although* the agent chooses a desirable level of the effort for the principal. In psychological equilibrium points, the agent's sense of guilty reduces the error probability from that in the subgame perfect equilibrium point of the modified inspector leadership game without psychological factors, and

*Graduate school of Humanities and Social Sciences, University of Tsukuba,1-1-1 Ten'no-dai, Ibaraki 305-8571, Japan; fukuzumi@social.tsukuba.ac.jp

the agent's impulse to deceive the principal raises the error probability from that in the subgame perfect equilibrium point of the modified inspector leadership game without psychological factors.

Keywords: inspection game, psychological game, sense of guilty, impulse to deceive, costly monitoring,

JEL classification numbers: C19, C72, J33.

1 Introduction

We formulate a modified inspector leadership game which consists of four stages. At the first stage of the game a principal (an inspector) decides a *critical output*. Given the *critical output*, an agent (an inspectee) chooses either a high effort e_H or a low effort e_L at the second stage of the game. Each effort e_l $(l = H, L)$ is a nonnegative real number. At the third stage an output is realized. The output is a random variable conditional on the effort level which was chosen by the agent at the previous stage. Finally, at the fourth stage of the game the principal observes the realized output but not the effort level which was chosen by the agent at the second stage of the game. If the realized output is below the *critical output* decided by the principal at the first stage of the game, then the principal conducts a *costly* investigation that provides the principal with information on the effort level which was chosen by the agent at the second stage. We assume that the investigation reveals *perfectly* the effort level chosen by the agent. In case that the agent chooses the low effort e_L and the principal conducts the investigation, the principal imposes a penalty on the agent, namely the principal gives the agent a low wage w_L. If the realized output is above the *critical output*, then the principal does not conduct the investigation into the effort level chosen by the agent and gives a higher wage w_H than w_L to the agent whether the agent chooses the high effort e_H or the low effort e_L. In case that the agent chooses the high effort e_H, the principal gives a higher wage w_H than w_L to the agent whether the principal conducts the invesigation or not.

The principal's payoff is defined as follows:

· If the principal conducts the investigation into the effort level which was chosen by the agent at the second stage, the principal's payoff is the output realized according to the effort level chosen by the agent minus the wage given to the agent minus the cost for conducting the investigation.

· If the principal does not conduct the investigation into the effort level which was chosen by the agent, the principal's payoff is the output realized according to the effort level chosen by the agent minus the wage given to the agent.

The payoff of the agent is the wage given by the principal minus the effort level chosen by the agent. In Section 2 we formulate this modified inspector leadership game and investigate the subgame perfect equilibrium point of this game.[1]

In Section 3 we introduce a psychological factor, the agent's sense of guilty, into our modified inspector leadership game and show that there are psychological equilibrium points such that the *critical output* chosen by the principal is smaller than that in the subgame perfect equilibrium point of our modified inspector leadership game in the previous section.[2]

In Section 4 we introduce a psychological factor, the agent's impulse to deceive the principal, into our modified inspector leadership game and show that there are psychological equilibrium points such that the *critical output*

---

[1]Useful techniques for material accountancy and data verification are developed in literatures of inspection games. Techniques developed for data verification are applicable to our modified inspector leadership game. For details of these useful techniques of inspection games, see Avenhause, Okada and Zamir (1991) and Avenhaus, Stengel and Zamir (2003).

[2]Many studies of experiments indicate that players are motivated by some psychological factors. For details of the progress of game theory with psychological factors, see Camerer (2003). Englemaier (2005) is a survey of behavioral game theoretic models about workers' moral hazard problems.

chosen by the principal is larger than that in the subgame perfect equilibrium point of our modified inspector leadership game in Section 2.[3]

2 The model without psychological factors

We consider a dynamic game with two players, P (principal) and A (agent). The game is described in Figure 1.

<Figure 1>

The game consists of four stages as following;

1. At the first stage of the game, player P chooses a probability $\alpha \in [0, 1]$ that player P monitors player A *ex post* by conducting an investigation that provides player P with information on an effort level chosen by player A. [4]

2. At the second stage of the game, player A given the probability α chooses his effort level e_l from a set $\{e_H, e_L\}$. Each effort level e_l is a nonnegative real number and $e_H > e_L$. We consider a behavior strategy given by the probability $q \in [0, 1]$ for choosing the low effort e_L.

[3]In Section 3 and in Section 4 we construct these models with psychological factors by the use of *psychological game theory*. Psychological game theory is proposed by Geanakoplos, Pearce and Stacchetti (1989) and Rabin (1993).

[4]$[0, 1]$ denotes an closed interval with end points 0 and 1. In the literature, we use similar notations. For example, (a, b) denotes an open interval and $(a, b]$ a semi closed interval with end points a and b.

3. At the third stage of the game, Nature picks up an output $y \in \mathbf{R}$. If player A has chosen e_l at the previous stage, the corresponding output y is realized according to a cumulative distribution function $F_l(y)$ which has the mean $\mu_l \in [0, \infty)$ where $l = L, H$, and $\mu_H > \mu_L$. Each distribution function F_l $(l = L, H)$ is absolutely continuous, has an inverse function F_l^{-1}, and has an identical variance with each other.

4. At the fourth stage, if an output y realized at the previous stage belongs to a set $Z_\alpha \equiv \{\ y \mid F_H(y) \leq \alpha\}$, player P conducts an investigation for the effort level chosen by player A. It costs a fixed amount of $c > 0$ unit of output for player P to conduct the investigation. After the investigation;

   - if player A has chosen e_H at the second stage of the game, then player P gives a fixed wage $w_H \in \mathbf{R}$ to player A,
   - if player A has chosen e_L at the second stage of the game, then player P gives a fixed wage $w_L \in \mathbf{R}$ to player A where $w_L < w_H$.

   If an output y realized at the previous stage does not belong to the set $Z_\alpha \equiv \{\ y \mid F_H(y) \leq \alpha\}$, then player P does not conduct the investigation and gives the fixed wage w_H to player A.

Payoff of each player

Let e_l be the effort level chosen by the agent at the second stage of the game. If the output y realized at the third stage of the game belongs to the set Z_α, each payoff of player P and of player A is given by $y - w_l$ and $w_l - e_l$, respectively. If the output y realized at third stage of the game does *not* belong to the set Z_α, each payoff of player P and of player A is given by $y - w_H$ and $w_H - e_l$, respectively. Let $z_\alpha \equiv F_H^{-1}(\alpha)$. Then the expected payoff of each player in our model is given by

$$\begin{aligned} Eu_A(\alpha, q) &= q\Big[\int_{z_\alpha}^{+\infty}(w_H - e_L)dF_L(y) + \int_{-\infty}^{z_\alpha}(w_L - e_L)dF_L(y)\Big] \\ &+ (1-q)\Big[\int_{z_\alpha}^{+\infty}(w_H - e_H)dF_H(y) + \int_{-\infty}^{z_\alpha}(w_H - e_H)dF_H(y)\Big], \end{aligned} \tag{2.1}$$

$$\begin{aligned} Eu_P(\alpha, q) &= q\Big[\int_{z_\alpha}^{+\infty}(y - w_H)dF_L(y) + \int_{-\infty}^{z_\alpha}(y - w_L - c)dF_L(y)\Big] \\ &+ (1-q)\Big[\int_{z_\alpha}^{+\infty}(y - w_H)dF_H(y) + \int_{-\infty}^{z_\alpha}(y - w_L - c)dF_H(y)\Big]. \end{aligned} \tag{2.2}$$

A market-imposed minimal expected payoff for player A is 0 and we assume that $w_L - e_L = 0$.

Relationship between hypothesis testing in statistics and our model

The null hypothesis H_0 in our model is that player A chooses the high effort e_H, and the alternative hypothesis H_1 in our model is that player A chooses the low effort e_L. The probability α chosen by player P is that of *the error of the first kind* in hypothesis testing. Namely, the value of α is the probability that the principal conducts the *costly* investigation *althouth* the agent chooses the high effort e_H.

We denote by $\beta \in [0, 1]$ the probability of *the error of the second kind* in hypothesis testing. Namely, the value of β is the probability that the player

P does *not* conduct the investigation into the effort level chosen by player A *although* player A chooses the low effort e_L. Moreover, we obtain a function $\beta = 1 - F_L(z_\alpha)$ where $z_\alpha \equiv F_H^{-1}(\alpha)$. The function $\beta(\alpha)$ fulfills $\beta(0) = 1$ and $\beta(1) = 0$.

Assumption 1. The function $\beta(\alpha) \in [0,1]^{[0,1]}$ is a differentiable, convex, and monotonically decreasing function.[5]

2.1 Analysis

The best response correspondence of player A

In order to investigate subgame perfect equilibrium points of our model, we consider the best response correspondence $q^*(\alpha)$ of player A to each $\alpha \in [0,1]$ chosen by player P. From (2.1) and the definition of $\beta(\alpha)$, the expected payoff of player A is given by

$$Eu_A(\alpha, q) = q\{(1 - \beta(\alpha))(w_H - w_L) + (e_H - e_L)\} + w_H - e_H. \quad (2.1.a)$$

Since we are interested in a case that for some $\alpha \in (0,1)$ player A has an incentive to choose the high effort e_H, we assume that $w_H - e_H > w_L - e_L$. In the following, we assume this inequality without further remark. Proofs of lemmas in this section are relegated to the Appendix A.

Lemma 2.1. Let $\alpha^* \equiv \beta^{-1}(1 - \frac{e_H - e_L}{w_H - w_L})$ where β^{-1} is an inverse function of

[5] $\beta(\alpha) \in [0,1]^{[0,1]}$ denotes a function $\beta(\alpha)$ on $[0,1]$ into $[0,1]$. In the following, we use similar notations. For example, $G(\alpha) \in \mathbf{R}^{[0,1]}$ denotes a function $G(\alpha)$ on $[0,1]$ into $\mathbf{R}$.

$\beta(\alpha)$. Then the best response correspondence of player A is given by

$$q^*(\alpha) = \begin{cases} 1 & \text{if } \alpha < \alpha^*, \\ [0,1] & \text{if } \alpha = \alpha^*, \\ 0 & \text{if } \alpha > \alpha^*. \end{cases}$$

The expected payoff of player P and the subgame perfect equilibrium points

Given the best response correspondence $q^*(\alpha)$ of player A, from (2.2) and the definition of $\beta(\alpha)$, the expected payoff of player P is given by

$$\begin{aligned} Eu_P(\alpha, q^*(\alpha)) &= (1 - q^*(\alpha))\{\mu_H - c\alpha - w_H\} \\ &+ q^*(\alpha)\{\beta(\alpha)(w_L + c - w_H) + \mu_L - w_L\} \qquad (2.2.a) \end{aligned}$$

The expected payoff $Eu_P(\alpha, q^*(\alpha))$ of player P has following properties.

Lemma 2.2.
(1) When $w_H - w_L > c$, $Eu_P(\alpha, q^*(\alpha))$ is an increasing function on $[0, \alpha^*)$ and a decreasing function on $(\alpha^*, 1]$.
(2) When $w_H - w_L \leq c$, $Eu_P(\alpha, q^*(\alpha))$ is a non-increasing function on $[0, \alpha^*)$ and a decreasing function on $(\alpha^*, 1]$.

Let $Eu_p(\alpha, 0) \equiv \mu_H - c\alpha - w_H$ and $Eu_p(\alpha, 1) \equiv \beta(\alpha)(w_L + c - w_H) + \mu_L - w_L$.

Lemma 2.3. Whenever $(\mu_H - w_H) - (\mu_L - w_L) > c$,
(1) $Eu_P(\alpha, 0) > Eu_P(\alpha, 1)$ for each $\alpha \in [0, 1]$,
(2) $Eu_P(0, 1) < Eu_P(\alpha^*, 0)$.

Due to the inequality $(\mu_H - w_H) - (\mu_L - w_L) > c$, player A must choose the high effort in the subgame perfect equilibrium point in our model. In the following, we assume the inequality $(\mu_H - w_H) - (\mu_L - w_L) > c$ without further remark. According to Lemma 2.2 and Lemma 2.3, graphs of the expected payoff $Eu_P(\alpha, q^*(\alpha))$ of player P are drawn at Figure 2.

<Figure 2>

Theorem 2.1. The subgame perfect equilibrium point (α^*, q^*) of our model without psychological factors is given by a pair of $\alpha^* = \beta^{-1}(1 - \frac{e_H - e_L}{w_H - w_L})$ and $q^* = 0$.

(Proof:) When $w_H - w_L > c$, an equality $q^*(\alpha) = 0$ must holds in the subgame perfect equilibrium point by part (1) of Lemma 2.3. Then by Lemma 2.1 the optimal strategy of player P is α^*. When $w_H - w_L \leq c$, an equality $q^*(\alpha) = 0$ must holds in the subgame perfect equilibrium point by part (1) and part (2) of Lemma 2.3. Then by Lemma 2.1 the optimal strategy of player P is α^*. ■

3 The model with a sense of guilty

We introduce a psychological factor, a sense of guilty of player A, into our model. Let $q'' \in [0, 1]$ be player A's belief about player P's belief about a behavior strategy $q \in [0, 1]$ which is a probability that player A chooses the low effort e_L. We call the belief $q'' \in [0, 1]$ the *second order belief* of player A.

Consider a situation where player A plays the low effort e_L and player P does *not* conduct an investigation into the effort level chosen by player A. In this situation the second belief q'' of player A is the smaller one, the more player A may feel guilty about his choosing the low effort.

Assumption 2. $g \in \mathbf{R}^{\mathrm{R}}$ is a differentiable and monotonically increasing function and fulfills that $g(0) = -k$ and $g(1) = 0$ where $k > 0$.

We add the function g to player A's payoff of this situation. Note that the function g is defined not only on $[0, 1]$ but also on the set of all real numbers $\mathbf{R}$, so that we can define an inverse function g^{-1} of g where the range of g^{-1} is the set of all real numbers $\mathbf{R}$. This change in our model is described at Figure 3.[6] Namely, the value of $-g(q'')$ captures the strength of the sense of guilty of player A with q''.

<Figure 3>

In the following of this paper an equilibrium concept to analyze our model with a psychological factor was given by Geanakoplos, Pearce and Stacchetti (1989) and Rabin (1993).

Definition 3.1. A *psychological equilibrium point* of our model with a psychological factor is a triplet $(\alpha^{**}, q^{**}, q'')$ such that
(1) the pair of (α^{**}, q^{**}) is the subgame perfect equilibrium point of our model with a psychological factor and
(2) $q^{**} = q''$. (*consistency*)

[6] Dufenberg (2002) proposed a trust game with a sense of guilty.

3.1 Analysis

The best response correspondence of player A

In order to investigate the psychological equilibrium point of our model with a sense of guilty of player A, we consider the best response correspondence $q^*(\alpha, q'')$ of player A to each pair $(\alpha, q'') \in [0,1]^2$. From (2.1) and Figure 3, the expected payoff $Eu_A(\alpha, q, q'')$ of player A with q'' is given by

$$\begin{aligned} Eu_A(\alpha, q, q'') &= q\{\beta(\alpha)\big(w_H - w_L + g(q'')\big) \\ &+ (w_L - w_H) + (e_H - e_L)\} + (w_H - e_H). \end{aligned} \quad (2.1.b)$$

Let $G(\alpha) \equiv g^{-1}\big(\frac{(w_H - w_L)+(e_L - e_H)}{\beta(\alpha)} - (w_H - w_L)\big)$ where g^{-1} is an inverse function of g and $\alpha \neq 1$.
Proofs of lemmas in this section are relegated to the Appendix B.

Lemma 3.1. The best response correspondence $q^*(\alpha, q'')$ of player A with q'' is given by

$$q^*(\alpha, q'') = \begin{cases} 1 & \text{if } q'' > G(\alpha), \\ [0,1] & \text{if } q'' = G(\alpha), \\ 0 & \text{if } q'' < G(\alpha) \text{ or } \alpha = 1. \end{cases}$$

We are interested in the minimum value of α that induces player A with *some* q'' to choose the high effort e_H. This minimum value of α is given by an equation $G(\alpha) = 0$ due to monotonicity of the function $G(\alpha)$.

Lemma 3.2. There is a number $\alpha_l \in [0, \alpha^*)$ such that $G(\alpha_l) = 0$ if and only if $e_H - e_L \geq k$.

Since $G(\alpha)$ is a monotonically increasing function of α, to each $\alpha \leq \alpha_l$ the optimal strategy of player A with each $q'' \in [0,1]$ is choosing the low effort e_L. According to Lemma 3.1 and Lemma 3.2, the best response correspondence $q^*(\alpha, q'')$ of player A with q'' is drawn at Figure 4.

<Figure 4>

The expected payoff of player P and the psychological equilibrium points

Lemma 3.3. Given the best response correspondence $q^*(\alpha, q'')$ of player A with a second order belief q'', the expected payoff $Eu_P(\alpha, q^*(\alpha, q''))$ of player P is given by

(1) $\beta(\alpha)(w_L + c - w_H) + \mu_L - w_L$

on a set $S_1 = \{(\alpha, q'') \in [0,1]^2 \mid 0 \leq \alpha < \alpha_l,\ 0 \leq q'' \leq 1\}$ and

on a set $S_2 = \{(\alpha, q'') \in [0,1]^2 \mid \alpha_l \leq \alpha < \alpha^*,\ q'' > G(\alpha)\}$.

(2) $\mu_H - c\alpha - w_H$

on a set $S_3 = \{(\alpha, q'') \in [0,1]^2 \mid \alpha_l < \alpha < \alpha^*,\ q'' < G(\alpha)\}$ and

on a set $S_4 = \{(\alpha, q'') \in [0,1]^2 \mid \alpha^* < \alpha < 1,\ 0 \leq q'' \leq 1\}$.

Let $k \leq e_H - e_L$. Then, for each $\bar{\alpha} \in [\alpha_l, \alpha^*]$, there is a second order belief $\bar{q''} \in [0,1]$ such that $\bar{q''} = G(\bar{\alpha})$. By Lemma 3.3 and monotonicity of $G(\alpha)$, for this $\bar{q''}$, the expected payoff of player P choosing $\alpha < \bar{\alpha}$ is given by $Eu_P(\alpha, q^*(\alpha, \bar{q''})) = \beta(\alpha)(w_L + c - w_H) + \mu_L - w_H$. Similarly, the expected payoff of player P choosing $\alpha > \bar{\alpha}$ is given by $Eu_P(\alpha, q^*(\alpha, \bar{q''})) = \mu_H - c\alpha - w_H$. The expected payoff $Eu_P(\alpha, q^*(\alpha, \bar{q''}))$ for the fixed $\bar{q''}$ is drawn at Figure 5. While the graph of $Eu_P(\alpha, q^*(\alpha, \bar{q''}))$ for the fixed $\bar{q''}$

jumps at a point $\bar{\alpha} \in [\alpha_l, \alpha^*]$, its shape of the graph is the same as the graph of $Eu_p(\alpha, q^*(\alpha))$ in Section 2 except the point that the graph jumps at.

<Figure 5>

Let $c < w_H - w_L$. Then $Eu_P(\alpha, 1)$ is a monotonically increasing function, and $Eu_P(\alpha, 0)$ is a monotonically decreasing function. Therefore, there is no incentive for player P to change $\bar{\alpha}$ to any other strategies. By Lemma 3.3, $q^*(\bar{\alpha}, \bar{q''}) = [0, 1]$, and by Lemma 2.3, $Eu_p(\alpha, 0) > Eu_p(\alpha, 1)$ for each $\alpha \in [0, 1]$. For $\bar{\alpha}$ to constitute player P's strategy of a subgame perfect equilibrium point, an equality $q^*(\alpha, q'') = 0$ must hold. By the consistency condition of Definition 3.1, an equality $q^{**} = q'' = 0$ holds in a psychological equilibrium point of our model. Hence it turns out that a triplet of $\bar{\alpha} \in [\alpha_l, \alpha^*]$, $q^{**} = 0$ and $q'' = 0$ is one of the psychological equilibrium points.

Let $S_p \equiv \{\alpha \in [0, 1] \mid \exists\ q'' \in [0, 1]$ such that $q'' = G(\alpha)\}$. We call a set $E_p \equiv \{(\alpha^{**}, q^{**}) \in [0, 1]^2 \mid (1)\ \alpha^{**} \in S_p,\ \ (2)$ a pair (α^{**}, q^{**}) is a subgame perfect equilibrium$\}$ a *possible set of psychological equilibrium points*. We use these concepts S_p and E_p, in the following cases of this Section 3 and in the next Section 4.

Case 1 ($k \leq e_H - e_L$, $c < w_H - w_L$)

Lemma 3.4. Each $\alpha \in [0, \alpha_l) \cup (\alpha^*, 1]$ is *not* the strategy of the psychological equilibrium point of player P.

Theorem 3.1. The possible set of psychological equilibrium points is given by $\{(\alpha^{**}, q^{**}, q'') \mid \alpha_l \leq \alpha^{**} \leq \alpha^*$ and $q^{**} = q'' = 0\}$.

(Proof:) Fix a strategy $\alpha \in [\alpha_l, \alpha^*]$ of player P. Since $G(\alpha)$ is a monotonically increasing and differentiable function on (α_l, α^*), there is a second order belief $q'' \in [0,1]$ such that $q'' = G(\alpha)$, so that $\alpha \in S_p$.

For this pair of α and q'', $q^*(\alpha, q'') = [0,1]$ by Lemma 3.3. For this α to constitute player P's strategy of a subgame perfect equilibrium point, an equality $q^*(\alpha, q'') = 0$ must hold, because by Lemma 2.3, $Eu_p(\alpha, 0) > Eu_p(\alpha, 1)$ for each $\alpha \in [0,1]$. By the consistency condition of Definition 3.1, $q^{**} = q'' = 0$ in a psychological equilibrium point of our model. Hence it turns out that a triplet of this α, $q^{**} = 0$ and $q'' = 0$ is one of the psychological equilibrium points.

By Lemma 3.4, each $\alpha \notin [\alpha_l, \alpha^*]$ is not the strategy of the psychological equilibrium point of player P. We obtain the result. ■

Case 2 ($k > e_H - e_L$, $c < w_H - w_L$)

Note that $S_1 = \phi$ and other sets S_i $(i = 2,3,4)$ in Lemma 3.3 are non-empty.

Lemma 3.5. Each $\alpha \in (\alpha^*, 1]$ is *not* the strategy of the psychological equilibrium point of player P.

Theorem 3.2. The possible set of psychological equilibrium points is given by $\{(\alpha^{**}, q^{**}, q'') \mid \ 0 \leq \alpha^{**} \leq \alpha^* \text{ and } q^{**} = q'' = 0\}$.

(Proof:) Fix a strategy $\alpha \in [0, \alpha^*]$ of player P. Since $k > e_H - e_L$, there is a second order belief $q'' \in (0,1]$ such that $q'' = G(\alpha)$, so that $\alpha \in S_p$.

For this α and q'', $q^*(\alpha, q'') = [0,1]$. For α to be player P's strategy of the subgame perfect equilibrium point, an equality $q^*(\alpha, q'') = 0$ must hold because $Eu_p(\alpha, 0) > Eu_p(\alpha, 1) > 0$ for each $\alpha \in [0,1]$ by Lemma 2.3. By the consistency condition of Definition 3.1, in a psychological equilibrium point of our model $q^{**} = q'' = 0$. Hence it turns out that a triplet of this α, $q^{**} = 0$ and $q'' = 0$ is one of the psychological equilibrium points.

By Lemma 3.5, each $\alpha \notin [0, \alpha^*]$ is not the strategy of the psychological equilibrium point of player P. We obtain the result. ■

Case 3 ($k \leq e_H - e_L$, $c > w_H - w_L$)

Note that each set S_i $(i = 1, 2, 3, 4)$ in Lemma 3.3 is non-empty.[7]

Lemma 3.6. Each $\alpha \in (0, \alpha_l) \cup (\alpha^*, 1]$ is *not* the strategy of the psychological equilibrium point of player P.

Theorem 3.3. The possible set of psychological equilibrium points is given by $\{(\alpha^{**}, q^{**}, q'') \mid \alpha_l \leq \alpha^{**} \leq \alpha^* \text{ and } q^{**} = q'' = 0\}$.

(Proof:) Fix strategy $\alpha \in [\alpha_l, \alpha^*]$ of player P. Then there is a second order belief $q'' = G(\alpha)$, so that $\alpha \in S_p$. For this α and q'', $q^*(\alpha, q'') = [0, 1]$. For α to be an optimal strategy for player P, $q^*(\alpha, q'') = 0$ must hold by part (1) of Lemma 2.4. By Lemma 3.6, each $\alpha \in (0, \alpha_l) \cup (\alpha^*, 1]$ is not player P's strategy of the psychological equilibrium point of player P. By part (2) of Lemma 2.4, $\alpha = 0$ is not player P's strategy of the psychological equilibrium point of player P. We obtain the result. ■

Case 4 ($k > e_H - e_L$, $c > w_H - w_L$)

Note that $S_1 = \phi$ and other sets S_i $(i = 2, 3, 4)$ in Lemma 3.3 are non-empty.

[7] We can deal with a case of $c = w_H - w_L$ in a similar way to the following Case 3 and Case 4.

Lemma 3.7. Each $\alpha \in (\alpha^*, 1]$ is not the strategy of the psychological equilibrium point of player P.

Theorem 3.4. The set of possible psychological equilibrium points is given by $\{(\alpha^{**}, q^{**}, q'') \mid \ 0 \leq \alpha^{**} \leq \alpha^* \text{ and } q^{**} = q'' = 0\}$.

(Proof:) Fix a strategy $\alpha \in [0, \alpha^*]$ of player P. Then there is a second order belief $q'' = G(\alpha)$, so that $\alpha \in S_p$. For this α and q'', $q^*(\alpha, q'') = [0, 1]$. For α to be an optimal strategy, $q^*(\alpha, q'') = 0$ must hold. By Lemma 3.7, each $\alpha \in (\alpha^*, 1]$ is not the strategy of the psychological equilibrium point of player P. We obtain the result. ■

4 The model with an impulse to deceive

We introduce a psychological factor, player A's impulse to deceive player P, into our model formulated in Section 2. Consider a situation where player A chooses the low effort e_L and player P does *not* conduct an investigation for the effort level chosen by player A. In this situation the second order belief q'' of player A for choosing the low effort e_L is the larger one, this situation gives the feeling of the more satisfaction to player A with an *impulse to deceive* player P.

Assumption 3. $g_d \in \mathbf{R}^{\mathbf{R}}$ is a differentiable and monotonically increasing function and fulfills that $g_d(0) = 0$ and $g_d(1) = k$ where $k > 0$.

We add the function g_d to player A's payoff of this situation. This change in our model is described at Figure 6.

<Figure 6>

We investigate the *psychological equilibrium point* in our model with player A's impulse to deceive.

The best response correspondence of player A

We replace the function g in the proof of Lemma 3.1 with g_d, so that obtain following Lemma 4.1. Proofs of lemmas in this section are relegated to the Appendix B. Let $G_d(\alpha) \equiv g_d^{-1}\big(\frac{(w_H - w_L)+(e_L - e_H)}{\beta(\alpha)} - (w_H - w_L)\big)$ where g_d^{-1} is an inverse function of g_d where $\alpha \neq 1$.

Lemma 4.1 The best response correspondence $q^*(\alpha, q'')$ of player A with q'' is given by

$$q^*(\alpha, q'') = \begin{cases} 1 & \text{if } q'' > G_d(\alpha), \\ [0,1] & \text{if } q'' = G_d(\alpha), \\ 0 & \text{if } q'' < G_d(\alpha) \text{ or } \alpha = 1. \end{cases}$$

We are interested in the maximum value of α which induces player A with *some* q'' to choose the high effort e_H. The maximum value of α is given by an equation $G_d(\alpha) = 1$ due to monotonicity of the function $G_d(\alpha)$.

Lemma 4.2.
(1) $G_d(\alpha^*) = 0$,
(2) There is a number $\alpha_h \in (\alpha^*, 1]$ such that $G_d(\alpha_h) = 1$ if and only if $k \leq e_H - e_L$.

According to Lemma 4.1 and Lemma 4.2, the best response correspondence $q^*(\alpha, q'')$ of player A with q'' is drawn at Figure 7.

<**Figure 7**>

The expected payoff of player P and the psychological equilibrium points

Lemma 4.3. Given the best response correspondence $q^*(\alpha, q'')$ of player A with a second order belief q'', the expected payoff $Eu_P(\alpha, q^*(\alpha, q''))$ of player P is given by

(1) $\beta(\alpha)(w_L + c - w_H) + \mu_L - w_L$

on a set $S_{1d} = \{(\alpha, q'') \in [0,1]^2 \mid 0 \leq \alpha < \alpha^*,\ 0 \leq q'' \leq 1\}$ and

on a set $S_{2d} = \{(\alpha, q'') \in [0,1]^2 \mid \alpha^* \leq \alpha < \alpha_h,\ q'' > G_d(\alpha)\}$.

(2) $\mu_H - c\alpha - w_H$

on a set $S_{3d} = \{(\alpha, q'') \in [0,1]^2 \mid \alpha^* < \alpha \leq \alpha_h,\ q'' < G_d(\alpha)\}$ and

on a set $S_{4d} = \{(\alpha, q'') \in [0,1]^2 \mid \alpha_h < \alpha \leq 1,\ 0 \leq q'' \leq 1\}$.

Let $c < w_H - w_L$. In order to investigate the possible set of psychological equilibrium points in our model with player A's impulse on deceit, we use the same logic as used in Case 1 and in Case 2 in the previous section. We omit the proof of the following theorem.

Theorem 4.1. Let $c < w_H - w_L$.

(1) When $k \leq e_H - e_L$, The possible set of psychological equilibrium points is given by $\{(\alpha^{**}, q^{**}, q'') \mid\ \alpha^* \leq \alpha^{**} \leq \alpha_h \text{ and } q^{**} = q'' = 0\}$.

(2) When $k > e_H - e_L$, The possible set of psychological equilibrium points is given by $\{(\alpha^{**}, q^{**}, q'') \mid \alpha^* \leq \alpha^{**} \leq 1 \text{ and } q^{**} = q'' = 0\}$.

In case that $c > w_H - w_L$ and $k \leq e_H - e_L$, ($k > e_H - e_L$), there is a subtle difference between the following analysis in this section and the corresponding analysis, Case 3 (Case 4, resp.), in the previous section.[8] If there is some $\alpha'_h \in (\alpha^*, \alpha_h)$ ($\alpha'_h \in (\alpha^*, 1)$, resp.) such that $\mu_H - c\alpha'_h - w_H = \beta(0)(w_L + c - w_H) + \mu_L - w_L$, then at each $\alpha \in (\alpha'_h, \alpha_h)$ ($(\alpha'_h, 1)$,resp.), player P gets a lower expected payoff than that at a point $\alpha = 0$. See Figure 8.

<Figure 8>

Each $\alpha \in (\alpha'_h, \alpha_h)$ ($\alpha \in (\alpha'_h, 1)$, resp.) is not player P's strategy of the subgame perfect equilibrium point for each $q'' \in [0, 1]$, so that is not player P's strategy of the psychological equilibrium point. Noting this fact, we obtain following theorems.

Theorem 4.2. Let $c > w_H - w_L$ and $k \leq e_H - e_L$.
If there is a number $\alpha'_h \in (\alpha^*, \alpha_h)$ such that $\alpha'_h = \frac{\mu_H - \mu_L - c}{c}$, then the possible set of psychological equilibrium points is given by $\{(\alpha^{**}, q^{**}, q'') \mid \alpha^* \leq \alpha^{**} \leq \alpha'_h \text{ and } q^{**} = q'' = 0\}$.
If $\alpha_h = \frac{\mu_H - \mu_L - c}{c}$, then the possible set of psychological equilibrium points is given by $\{(\alpha^{**}, q^{**}, q'') \mid \alpha^* \leq \alpha^{**} \leq \alpha'_h \text{ and } q^{**} = q'' = 0\} \cup \{(0, 1, 1)\}$.
Otherwise, the possible set of psychological equilibrium points is given by $\{(\alpha^{**}, q^{**}, q'') \mid \alpha^* \leq \alpha^{**} \leq \alpha_h \text{ and } q^{**} = q'' = 0\}$.

(Proof:) An equality $\mu_H - c\alpha'_h - w_H = \beta(0)(w_L + c - w_H) + \mu_L - w_L$ is equivalent to $\alpha'_h = \frac{\mu_H - \mu_L - c}{c}$. Assume that $\alpha'_h \in (\alpha^*, \alpha_h)$. By Lemma 4.1 and Lemma 4.2, the best response correspondence $q^*(\alpha, q'')$ of player A with $q'' < G_d(\alpha)$ is given by $q^*(\alpha, q'') = 0$ for each $\alpha \in (\alpha'_h, \alpha_h)$. Then, on $(\alpha'_h, 1]$, the expected payoff $Eu_p(\alpha, q^*(\alpha, q''))$ of player P is smaller than that of $\alpha = 0$, namely

[8] We can deal with a case of $c = w_H - w_L$ in a similar way to the following.

$\mu_H - c\alpha - w_H < Eu_p(0, q^*(\alpha, q'')) = \beta(0)(w_L + c - w_H) + \mu_L - w_L$, because the function $\mu_H - c\alpha - w_H$ is monotonically decreasing. Hence we have a following claim.

Claim 1: Each $\alpha \in (\alpha'_h, 1]$ is *not* a player P's strategy of psychological equilibrium points of our model with an impulse of deceive.

For each $\alpha \in [0, \alpha^*)$, $Eu_p(\alpha, 1) < Eu_p(\alpha^*, 0)$ by part (2) of Lemma 2.3 and monotonicity of both $Eu_p(\alpha, 1)$ and $Eu_p(\alpha, 0)$. Hence, each $\alpha \in (0, \alpha^*)$ is not the optimal strategy for player P given player A's best responce correspondence. Hence we have a following claim.

Claim 2: Each $\alpha \in (0, \alpha^*)$ is *not* a player P's strategy of psychological equilibrium points.

Fix a strategy $\alpha \in [\alpha^*, \alpha'_h]$ of player P. Then there is a second order belief $q'' = G(\alpha)$, so that $\alpha \in S_p$. For this α and q'', $q^*(\alpha, q'') = [0, 1]$. For α to be an optimal strategy, $q^*(\alpha, q'') = 0$ must hold by part (1) of Lemma 2.3. By Claim 1 and Claim 2, each $\alpha \in (0, \alpha^*) \cup (\alpha'_h, 1]$ is not the strategy of the psychological equilibrium point of player P.

If $\alpha_h = \frac{\mu_H - \mu_L - c}{c}$, an equality $Eu_p(0, 1) = Eu_p(\alpha_h, 0)$ holds, so that obviously a point $(\alpha^{**}, q^{**}, q'') = (0, 1, 1)$ is also the psychological equilibrium point.

Assume that there is no $\alpha'_h \in (\alpha^*, \alpha_h)$ such that $\mu_H - c\alpha'_h - w_H = \beta(0)(w_L + c - w_H) + \mu_L - w_L$ and $\alpha_h \neq \frac{\mu_H - \mu_L - c}{c}$. Then each $\alpha \in (\alpha_h, 1]$ is not optimal starategy of player P by Lemma 4.3 and monotonicity of $Eu_p(\alpha, 1)$.

Each $\alpha \in [0, \alpha^*)$ is not optimal strategy of player P by part (2) of Lemma 2.3. Fix a strategy $\alpha \in [\alpha^*, \alpha_h]$ of player P. Then there is a second order belief $q'' = G(\alpha)$, so that $\alpha \in S_p$. For this α and q'', $q^*(\alpha, q'') = [0, 1]$. For α to be an optimal strategy, $q^*(\alpha, q'') = 0$ must hold.

We have thus proved the theorem. ■

Theorem 4.3. Let $c > w_H - w_L$ and $k > e_H - e_L$.
If there is a number $\alpha'_h \in (\alpha^*, 1)$ such that $\alpha'_h = \frac{\mu_H - \mu_L - c}{c}$, then the possible set of psychological equilibrium points is given by $\{(\alpha^{**}, q^{**}, q'') \mid \alpha^* \leq \alpha^{**} \leq \alpha'_h \text{ and } q^{**} = q'' = 0\}$.
If $\frac{\mu_H - \mu_L - c}{c} = 1$, then the possible set of psychological equilibrium points is given by $\{(\alpha^{**}, q^{**}, q'') \mid \alpha^* \leq \alpha^{**} \leq 1 \text{ and } q^{**} = q'' = 0\} \cup \{(0, 1, 1)\}$.
Otherwise, the possible set of psychological equilibrium points is given by $\{(\alpha^{**}, q^{**}, q'') \mid \alpha^* \leq \alpha^{**} \leq 1 \text{ and } q^{**} = q'' = 0\}$.

(Proof:) An equality $\mu_H - c\alpha'_h - w_H = \beta(0)(w_L + c - w_H) + \mu_L - w_L$ is equivalent to $\alpha'_h = \frac{\mu_H - \mu_L - c}{c}$. Assume that $\alpha'_h \in (\alpha^*, 1]$. By Lemma 4.1 and Lemma 4.2, the best response correspondence $q^*(\alpha, q'')$ of player A with $q'' < G_d(\alpha)$ is given by $q^*(\alpha, q'') = 0$ for each $\alpha \in (\alpha'_h, 1]$. Then, on $(\alpha'_h, 1]$, the expected payoff $Eu_p(\alpha, q^*(\alpha, q''))$ of player P is smaller than that of $\alpha = 0$, namely $\mu_H - c\alpha - w_H < Eu_p(0, q^*(\alpha, q'')) = \beta(0)(w_L + c - w_H) + \mu_L - w_L$, because the function $\mu_H - c\alpha - w_H$ is monotonically decreasing. Hence we have a following claim.

Claim 3: Each $\alpha \in (\alpha'_h, 1]$ is *not* a player P's strategy of psychological equilibrium points of our model with an impulse of deceive.

For each $\alpha \in (0, \alpha^*)$, $Eu_p(\alpha, 0) > Eu_p(\alpha, 1)$ by part (2) of Lemma 2.3 and monotonicity of both $Eu_p(\alpha, 0)$ and $Eu_p(\alpha, 1)$. Each $\alpha \in (0, \alpha^*)$ is not the optimal strategy for player P given player A's best response correspondence. Hence we have a following claim.

Claim 4: Each $\alpha \in [0, \alpha^*)$ is *not* a player P's strategy of psychological equilibrium points.

Fix a strategy $\alpha \in [\alpha^*, \alpha'_h]$ of player P. Then there is a second order belief $q'' = G(\alpha)$, so that $\alpha \in S_p$. For this α and q'', $q^*(\alpha, q'') = [0, 1]$. For α to be an optimal strategy, $q^*(\alpha, q'') = 0$ must hold. By Claim 3 and Claim 4, each $\alpha \in [0, \alpha^*) \cup (\alpha'_h, 1]$ is not the strategy of the psychological equilibrium point of player P.

If $\frac{\mu_H - \mu_L - c}{c} = 1$, an equality $Eu_p(0, 1) = Eu_p(1, 0)$ holds, so that obviously a point $(\alpha^{**}, q^{**}, q'') = (0, 1, 1)$ is also the psychological equilibrium point.

Assume that there is no $\alpha'_h \in (\alpha^*, 1]$ such that $\mu_H - c\alpha'_h - w_H = \beta(0)(w_L + c - w_H) + \mu_L - w_L$. Each $\alpha \in [0, \alpha^*)$ is not optimal strategy of player P by part (2) of Lemma 2.3. Fix a strategy $\alpha \in [\alpha^*, 1]$ of player P. Then there is a second order belief $q'' = G(\alpha)$, so that $\alpha \in S_p$. For this α and q'', $q^*(\alpha, q'') = [0, 1]$. For α to be an optimal strategy, $q^*(\alpha, q'') = 0$ must hold.

We have thus proved the theorem. ■

5 Concluding remarks

We show that a possible set of psychological equilibrium points constitutes a *connected* set, $[\alpha_l, \alpha^*]$, $[\alpha^*, \alpha_h]$ etc. Avenhaus, Okada and Zamir (1991) shows that in a inspector leadership game with incomplete information about player A's payoff, the inspector's payoff function given player A's best response becomes a *continuous* function of α. Our model with psychological factors gives a psychological foundation to incomplete information about player A's payoff.

In psychological terms g, g_d, of our model, player A does *not* take account of a difference between player P's payoff and player A' payoff, moreover we do not introduce a psychological factor of *fairness* between players. Each player with a psychological factor of *fairness* behaves *reciprocally* toward the opponent player. Behavioral game theories with players who have *reciprocal factors* in psychological terms have good predictions in experiments.[9] Falk and Fischbacher (2006) gives explanations for some anomalous behaviors ovserved in experiments by the use of their theory of reciprocity. With a *reciprocal factor*, our modified inspection game may have a good prediction of experiments.

We assume *fixed* wages for player A. Our results may change drastically if we consider more flexible systems of wages. Then we need to consider an optimization with respect to both the error probability α and the wage level. This is a challenging and interesting work. Dye (1986) and Kanodia (1985) study more flexible systems of wage than that of our models, but they do

[9] Dufwenberg and Kirchsteiger (2004) and Falk and Fischbacher (2006) propose theories of reciprocity.

not introduce psychological factors into their models.

Appendix A

This appendix A provides proofs of lemmas in Section 2.

Proof of Lemma 2.1. From (2.1.a), if $(1-\beta(\alpha))(w_H-w_L)+(e_H-e_L)>0$ then $q^*(\alpha)=1$. Since $w_H-w_L>0$, $\beta(\alpha)>1-\frac{e_H-e_L}{w_H-w_L}$. By Assumption 1, the function $\beta(\alpha)$ has an inverse function $\beta^{-1}\in[0,1]^{[0,1]}$ which is decreasing. Due to $w_H-e_H>w_L-e_L=0$, $\frac{e_H-e_L}{w_H-w_L}\in(0,1)$ and $\beta^{-1}(1-\frac{e_H-e_L}{w_H-w_L})$ is well defined. Hence we obtain $\alpha<\beta^{-1}(1-\frac{e_H-e_L}{w_H-w_L})$.

From (2.1.a) and the definition of $\beta(\alpha)$, if $(1-\beta(\alpha))(w_H-w_L)=0$, then $q^*(\alpha)\in[0,1]$.

If $(1-\beta(\alpha))(w_H-w_L)\leq 0$, then $q^*(\alpha)=0$. Since $w_H-w_L>0$, $\beta(\alpha)>1-\frac{e_H-e_L}{w_H-w_L}$. By Assumption 1, the function $\beta(\alpha)$ has an inverse function $\beta^{-1}\in[0,1]^{[0,1]}$ which is non-increasing. Due to $w_H-e_H>w_L-e_L$ and $e_H>e_L$, $\frac{e_H-e_L}{w_H-w_L}\in(0,1)$ and $\beta^{-1}(1-\frac{e_H-e_L}{w_H-w_L})$ is well defined. Hence we obtain $\alpha>\beta^{-1}(1-\frac{e_H-e_L}{w_H-w_L})$. ■

Proof of Lemma 2.2. (1) By Lemma 2.1 and (2.2.a), $Eu_P(\alpha,q^*(\alpha))=\mu_L+\beta(\alpha)(w_L+c-w_H)-w_L$ for each $\alpha\in[0,\alpha^*)$. Since $w_H-w_L>c$ and $\beta(\alpha)$ is a decreasing function of α, $Eu_P(\alpha,q^*(\alpha))$ is an increasing function on $[0,\alpha^*)$. By Lemma 2.1, $Eu_P(\alpha,q^*(\alpha))=\mu_H-c\alpha-w_H$ for each $(\alpha^*,0]$. Since $c>0$, we obtain the result.
(2) By Lemma 2.1 and (2.2.a), $Eu_P(\alpha,q^*(\alpha))=\mu_L+\beta(\alpha)(w_L+c-w_H)-w_L$ for each $\alpha\in[0,\alpha^*)$. Since $w_H-w_L\leq c$ and $\beta(\alpha)$ is a monotonically decreasing function of α, $Eu_P(\alpha,q^*(\alpha))$ is a non-increasing function on $[0,\alpha^*)$.

By Lemma 2.1, $Eu_P(\alpha, q^*(\alpha)) = \mu_H - c\alpha - w_H$ for each $(\alpha^*, 0]$. Since $c > 0$, we obtain the result. ■

Proof of Lemma 2.3. (1) $Eu_P(\alpha, 0) > Eu_P(\alpha, 1)$ is $\mu_H - c\alpha - w_H > \mu_L + \beta(\alpha)(w_L + c - w_H) - w_L$ from (2.2). This inequality becomes $\mu_H - \mu_L + (\beta(\alpha) - 1)(w_H - w_L) > (\alpha + \beta(\alpha))c$. By the definition of $\beta(\alpha)$ and a fact that $\alpha + \beta(\alpha) < 1$, for each $\alpha \in [0, 1]$ $\mu_H - \mu_L + (\beta(\alpha) - 1)(w_H - w_L) \geq \mu_H - \mu_L - w_H + w_L = (\mu_H - w_H) - (\mu_L - w_L) > c > (\alpha + \beta(\alpha))c$.
(2) Suppose, by contradiction, $Eu_P(0, 1) \geq Eu_P(\alpha^*, 0)$. Then there is a number $\alpha' \in [0, \alpha^*]$ such that $Eu_P(0, 1) = Eu_P(\alpha', 0)$ by part (2) of Lemma 2.2. We have $\alpha' = \frac{(\mu_H - w_H) - (\mu_L - w_L)}{c} + 1$. Since $(\mu_H - w_H) - (\mu_L - w_L) > 0$, a contradiction. ■

Appendix B

This appendix B provides proofs of lemmas in Section 3 and in Section 4.

Proof of Lemma 3.1. From (2.1.b), if an inequality $\beta(\alpha)(w_H - w_L + g(q'')) + (w_L - w_H) + (e_H - e_L) > 0$ holds, then $q^*(\alpha, q'') = 1$. Let $\beta(\alpha) \neq 0$. Then the inequality becomes $g(q'') > \frac{(w_H - w_L) + (e_L - e_H)}{\beta(\alpha)} - (w_H - w_L)$. Since g has a monotonically increasing inverse function $g^{-1} \in \mathbf{R}^{\mathbf{R}}$, $q'' > g^{-1}\big(\frac{(w_H - w_L) + (e_L - e_H)}{\beta(\alpha)} - (w_H - w_L)\big)$. A function $G(\alpha) \equiv g^{-1}(\frac{(w_H - w_L) + (e_L - e_H)}{\beta(\alpha)} - (w_H - w_L))$ is a monotonically increasing function of α on $[0, 1]$. This follows from facts that;

(1) $\beta(\alpha)$ is a monotonically increasing function of α,

(2) $w_H - e_H > w_L - e_L$,

(3) g^{-1} is a monotonically increasing function.

Therefore, if $q'' > G(\alpha)$ then $q^*(\alpha, q'') = 1$.

From (2.1.b), if an inequality $\beta(\alpha)\big(w_H - w_L + g(q'')\big) + (w_L - w_H) + (e_H - e_L) < 0$ holds, then $q^*(\alpha, q'') = 0$. Let $\beta(\alpha) \neq 0$. Then the inequality becomes $g(q'') < \frac{(w_H-w_L)+(e_L-e_H)}{\beta(\alpha)} - (w_H - w_L)$. Since g has the monotonically increasing inverse function $g^{-1} \in \mathbf{R}^{\mathbf{R}}$, $q'' < g^{-1}\big(\frac{(w_H-w_L)+(e_L-e_H)}{\beta(\alpha)} - (w_H-w_L)\big)$. The function $G(\alpha) \equiv g^{-1}\big(\frac{(w_H-w_L)+(e_L-e_H)}{\beta(\alpha)} - (w_H - w_L)\big)$ is a monotonically increasing function of α on $[0, 1]$, so that if $q'' < G(\alpha)$ then $q^*(\alpha, q'') = 0$.

From (2.1.b), if an equality $\beta(\alpha)\big(w_H - w_L + g(q'')\big) + (w_L - w_H) + (e_H - e_L) = 0$ holds, then player A randomly chooses his effort level. Let $\beta(\alpha) \neq 0$. Then the equality becomes $g(q'') = \frac{(w_H-w_L)+(e_L-e_H)}{\beta(\alpha)} - (w_H - w_L)$. Since g has the monotonically increasing inverse function $g^{-1} \in \mathbf{R}^{\mathbf{R}}$, $q'' = g^{-1}\big(\frac{(w_H-w_L)+(e_L-e_H)}{\beta(\alpha)} - (w_H-w_L)\big)$. The function $G(\alpha) \equiv g^{-1}\big(\frac{(w_H-w_L)+(e_L-e_H)}{\beta(\alpha)} - (w_H - w_L)\big)$ is a monotonically increasing function of α on $[0, 1]$, so that if $q'' = G(\alpha)$ then player A randomly chooses his effort level.

When $\beta(\alpha) = 0$, that is $\alpha = 1$, $\beta(\alpha)\big(w_H - w_L + g(q'')\big) + (w_L - w_H) + (e_H - e_L) = (w_L - w_H) + (e_H - e_L) < 0$ due to $w_H - e_H > w_L - e_L$. Then $q^*(1, q'') = 0$ for each $q'' \in [0, 1]$. ■

Proof of Lemma 3.2. Since $\beta(0) = 1$, $G(0) = g^{-1}(e_L - e_H) \leq 0$. By Assumption 2, $e_L - e_H < -k$ if and only if $G(0) < 0$. $G(\alpha) \in \mathbf{R}^{(0,1]}$ is a differentiable and monotonically increasing function due to Assumption 1 and Assumption 2. Since $\lim\limits_{\alpha \to +1} G(\alpha) = +\infty$, $e_L - e_H < -k$ if and only if there is a number $\alpha_l \in (0, 1)$ such that $G(\alpha_l) = 0$.

Since $G(\alpha^*) = G\big(1 - \frac{e_H-e_L}{w_H-w_L}\big) = 0$ and $G(\alpha) \in \mathbf{R}^{(0,1]}$ is monotonically increasing, we obtain $\alpha_l < \alpha^*$. ■

Proof of Lemma 3.3. (1) Since $G(\alpha)$ is monotonically increasing, $q'' \geq 0 \geq G(\alpha)$ for each point $(\alpha, q'') \in S_1$. Hence $q^*(\alpha, q'') = 1$ for each point $(\alpha, q'') \in S_1$ by Lemma 3.1. From (2.2.a), the expected payoff of player P is given by $Eu_p(\alpha, 1) = \beta(\alpha)(w_L + c - w_H) + \mu_L - w_L$ for each point $(\alpha, q'') \in S_1$.

For each point $(\alpha, q'') \in S_2$, $q^*(\alpha, q'') = 1$ by Lemma 3.1. From (2.2.a), the expected payoff of player P is given by $Eu_p(\alpha, 1) = \beta(\alpha)(w_L + c - w_H) + \mu_L - w_L$ for each point $(\alpha, q'') \in S_2$.

(2) For each point $(\alpha, q'') \in S_3$, $q^*(\alpha, q'') = 0$ by Lemma 3.1. From (2.2.a), the expected payoff of player P is given by $\mu_H - c\alpha - w_H$ for each point $(\alpha, q'') \in S_3$.

Since $G(\alpha)$ is monotonically increasing, $q'' < G(\alpha)$ for each point $(\alpha, q'') \in S_4$. Hence $q^*(\alpha, q'') = 0$ for each point $(\alpha, q'') \in S_4$ by Lemma 3.1. From (2.2.a), the expected payoff of player P is given by $\mu_H - c\alpha - w_H$ for each point $(\alpha, q'') \in S_4$. ■

Proof of Lemma 3.4. By Lemma 3.3, $q^*(\alpha, q'') = 1$ for each $\alpha \in [0, \alpha_l]$ and each $q'' \in [0, 1]$, so that $Eu_P(\alpha, q^*(\alpha, q'')) = Eu_P(\alpha, 1) = \beta(\alpha)(w_L + c - w_H) + \mu_L - w_L$ for each $\alpha \in [0, \alpha_l)$ and each $q'' \in [0, 1]$.
By Assumption 1 and an inequality $c < w_H - w_L$, the expected payoff function $Eu_P(\alpha, q^*(\alpha, q'')) = \beta(\alpha)(w_L + c - w_H) + \mu_L - w_L$ is monotonically increasing in $\alpha \in [0, \alpha_l)$, so that $\alpha \in [0, \alpha)$ is not a strategy of the subgame perfect equilibrium point of our model with a sense of guilty of the agent. Hence each $\alpha \in [0, \alpha_l)$ is not the psychological equilibrium point of our model.

By Lemma 3.3 and the fact $c > 0$, for each $\alpha \in (\alpha^*, 1]$, $Eu_P(\alpha, q^*(\alpha, q'')) = Eu_P(\alpha, 0) = \mu_H - c\alpha - w_H$ which is monotonically decreasing in $\alpha \in (\alpha^*, 1]$, so that $\alpha \in (\alpha^*, 1]$ is not a strategy of the subgame perfect equilibrium point of our model with a sense of guilty of the agent. Thus, each $\alpha \in (\alpha^*, 1]$ is not the psychological equilibrium point of our model. ■

Proof of Lemma 3.5. Since Lemma 3.3 and the consistency condition of Definition 1, $Eu_P(\alpha, q^*(\alpha, q'')) = Eu_P(\alpha, 0) = \mu_H - c\alpha - w_H$ for each $\alpha \in (\alpha^*, 1]$. This function $Eu_P(\alpha, q^*(\alpha, 0)$ is decreasing on $(\alpha^*, 1]$ due to $c > 0$, so that each $\alpha \in (\alpha^*, 1]$ is not player P's strategy of the subgame perfect equilibrium strategy. Hence each $\alpha \in (\alpha^*, 1]$ is not player P's strategy of the psychological equilibrium point of our model. ■

Proof of Lemma 3.6. By Lemma 3.3 $Eu_P(\alpha, q^*(\alpha, q'')) = Eu_P(\alpha, 1) = \beta(\alpha)(w_L + c - w_H) + \mu_L - w_L$ for each $\alpha \in [0, \alpha_l)$. By Assumption 1, if $c > w_H - w_L$ then the expected payoff function $Eu_P(\alpha, q^*(\alpha, q'')) = \beta(\alpha)(w_L + c - w_H) + \mu_L - w_L$ is monotonically decreasing in $\alpha \in [0, \alpha_l)$, so that each $\alpha \in (0, \alpha_l)$ is not a strategy of the subgame perfect equilibrium point. Hence each $\alpha \in (0, \alpha_l)$ is not a strategy of the psychological equilibrium point.

By Lemma 3.3, for each $\alpha \in (\alpha^*, 1]$, $Eu_P(\alpha, q^*(\alpha, q'')) = Eu_P(\alpha, 0) = \mu_H - c\alpha - w_H$ which is monotonically decreasing in $\alpha \in (\alpha^*, 1]$ due to $c > 0$. Thus, each $\alpha \in (\alpha^*, 1]$ is not a strategy of the psychological equilibrium point of our model. ■

Proof of Lemma 3.7. By Lemma 3.3, $Eu_P(\alpha, q^*(\alpha, q'')) = Eu_P(\alpha, 1) = \mu_H - c\alpha - w_H$ for each $\alpha \in (\alpha^*, 1]$. This expected payoff function $Eu_P(\alpha, q^*(\alpha, 0)$ of player P is monotonically decreasing in $\alpha \in (\alpha^*, 1]$ due to $c > 0$. Thus, each $\alpha \in (\alpha^*, 1)$ is not player P's strategy of the psychological equilibrium point of our model. ■

Proof of Lemma 4.2. (1) Let $G_d(\alpha) = 0$. Then $g_d(0) = \frac{(w_H - w_L) + (e_L - e_H)}{\beta(\alpha)} - (w_H - w_L) = 0$ by Assumption 3. We obtain $\alpha = \beta^{-1}\big(1 - \frac{e_H - e_L}{w_H - w_L}\big) = \alpha^*$.

(2) Let $G(\alpha) = 1$. Then $\frac{(w_H - w_L) + (e_L - e_H)}{\beta(\alpha)} - (w_H - w_L) = k$ by Assumption 3, so that $\beta(\alpha) = \frac{(w_H - w_L) + (e_L - e_H)}{w_H - w_L + k}$. By the definition of $\beta(\alpha)$, if $k < e_H - e_L$, there is a unique number $\alpha \leq 1$ such that $\beta(\alpha) = \frac{(w_H - w_L) + (e_L - e_H)}{w_H - w_L + k}$. Let the number α be α_h. By Assumption 2, $\alpha - \alpha^* = \beta^{-1}\big(\frac{w_H - w_L + e_L - e_H}{w_H - w_L + k}\big) - \beta^{-1}\big(\frac{w_H - w_L + e_L - e_H}{w_H - w_L}\big) \geq 0$. ■

Proof of Lemma 4.3. By Lemma 4.1, $q^*(\alpha, q'') = 1$ on S_{1d} and S_{2d}. From (2.2.a), the expected payoff of player P is given by $Eu_p(\alpha, 1) = \beta(\alpha)(w_L + c - w_H) + \mu_L - w_L$.

By Lemma 4.2, $q^*(\alpha, q'') = 0$ on S_{3d} and S_{4d}. From (2.2.a), the expected payoff of player P is given by $\mu_H - c\alpha - w_H$. ■

References

1. Avenhaus, R., and A. Okada (1992) "Statistical criteria for sequential inspector-leadership games," *Journal of the Operations Research Society of Japan* **35**, 134–151.

2. Avenhaus, R., A. Okada, and S. Zamir (1991) "Inspector leadership with incomplete information," in *Game Equilibrium Models* (eds. by R. Selten). Springer Berlin.

3. Avenhaus, R., B.-V. Stengel, and S. Zamir (2002) "Inspection games," in *Handbook of Game Theory* (eds. by R.-J. Aumann and S. Hart). 1948–1987. Elsevier Science B.V.

4. Camerer, C,-F. (2003) *Behavioral Game Theory*, Princeton University Press P.N.

5. Dye, R,-A. (1986) "Optimal monitoring policies in agencies," *Rand Journal of Economics* **17**, 339–350.

6. Dufwenberg, M., and G. Kirchsteiger (2004) "A theory of sequential reciprocity," *Games and Economic Behavior* **47**, 268–298.

7. Dufwenberg, M. (2002) "Marital investments, time consistency and emotions," *Journal of Economic Behavior and Organization* **48**, 57–69.

8. Englmaier, F. (2005) "Moral hazard, contracts and social preference," in *Psychology, Rationality and Economic Behavior* (eds. by B. Agarwal and A. Vercelli). Palgrave Macmillan H.B.H.

9. Falk, A., and U. Fischbacher (2006) "A theory of reciprocity," *Games and Economic Behavior* **54**, 293–315.

10. Geanakoplos, J., D. Pearce, and E. Stacchetti (1989) "Psychological games and sequential rationality," *Games and Economic Behavior* **1**, 60–79.

11. Kanodia, C,-S. (1985) "Stochastic monitoring and moral hazard," *Journal of Accounting Research* **23**, 175–193.

12. Rabin, M. (1993) "Incorporating fairness into game theory and economics," *American Economic Review* **83**, 1281–1302.

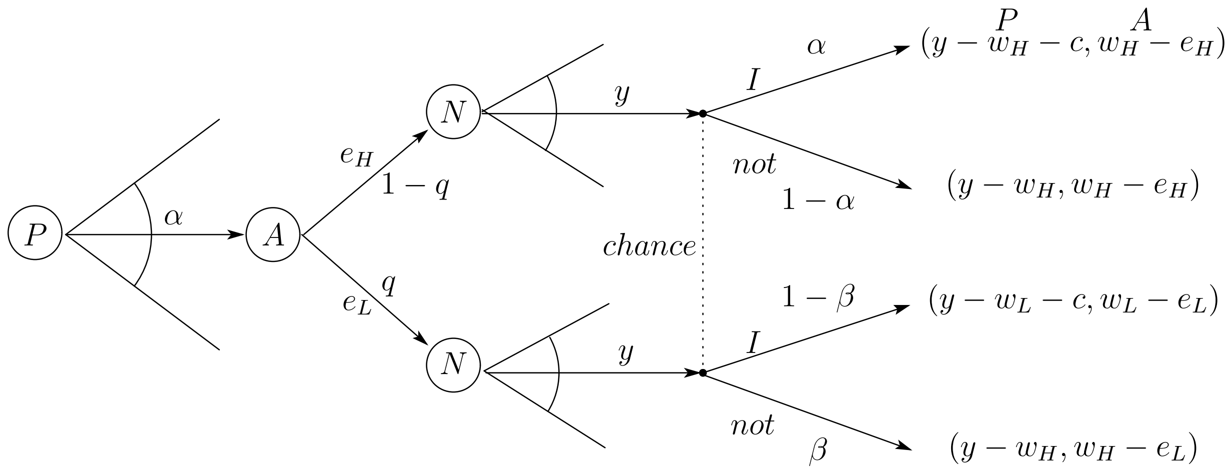

Figure 1: The model without psychological factors.

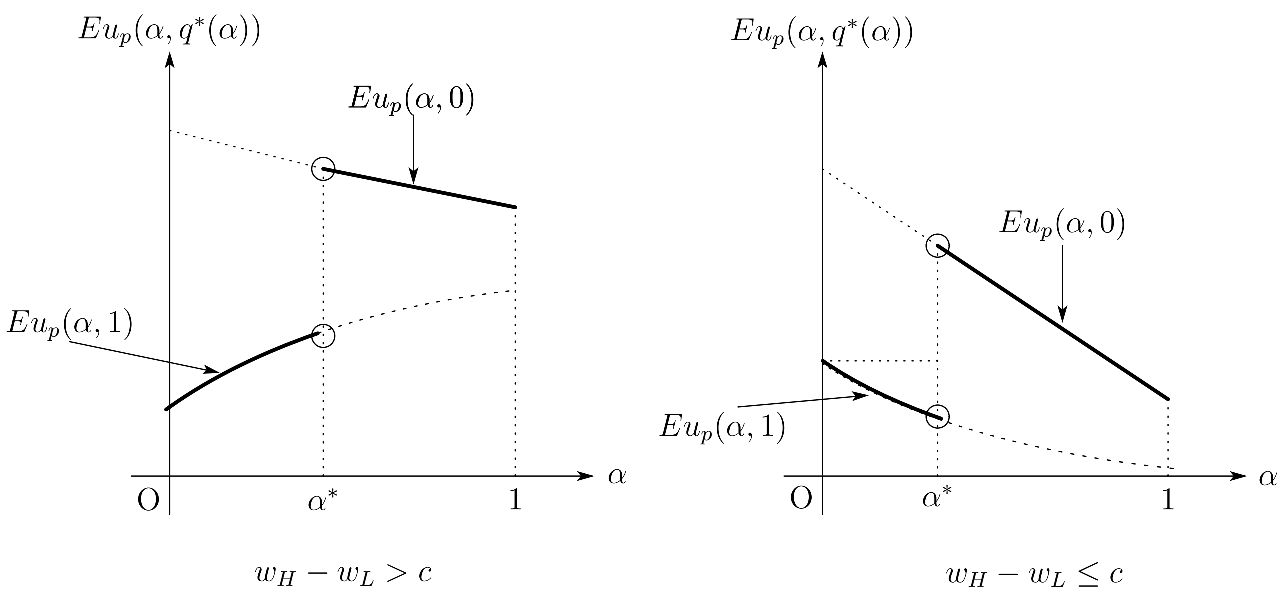

Figure 2: The expected payoff $Eu_p(\alpha, q^*(\alpha))$ of player P.

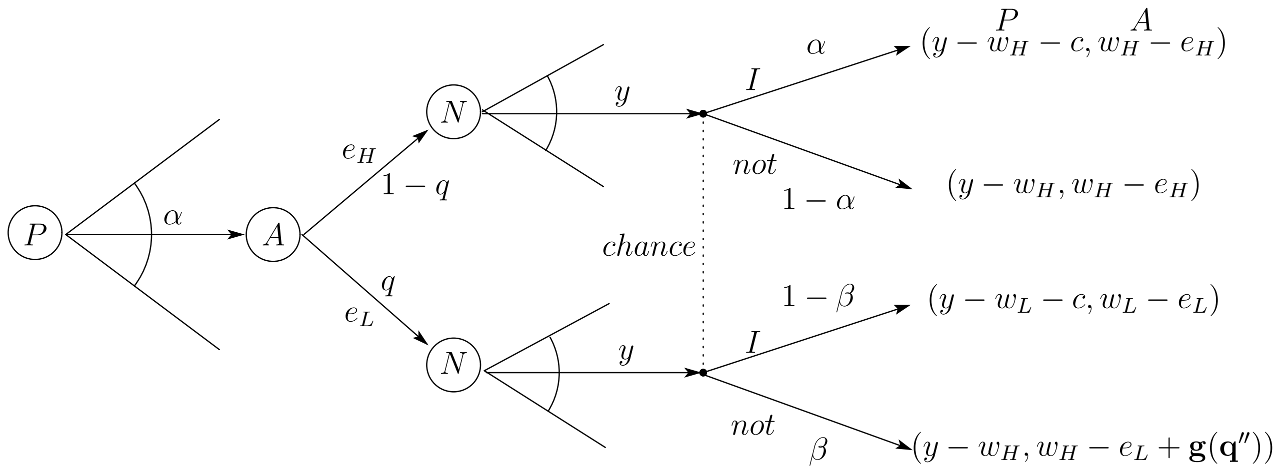

Figure 3: The model with a sense of guilty of player A.

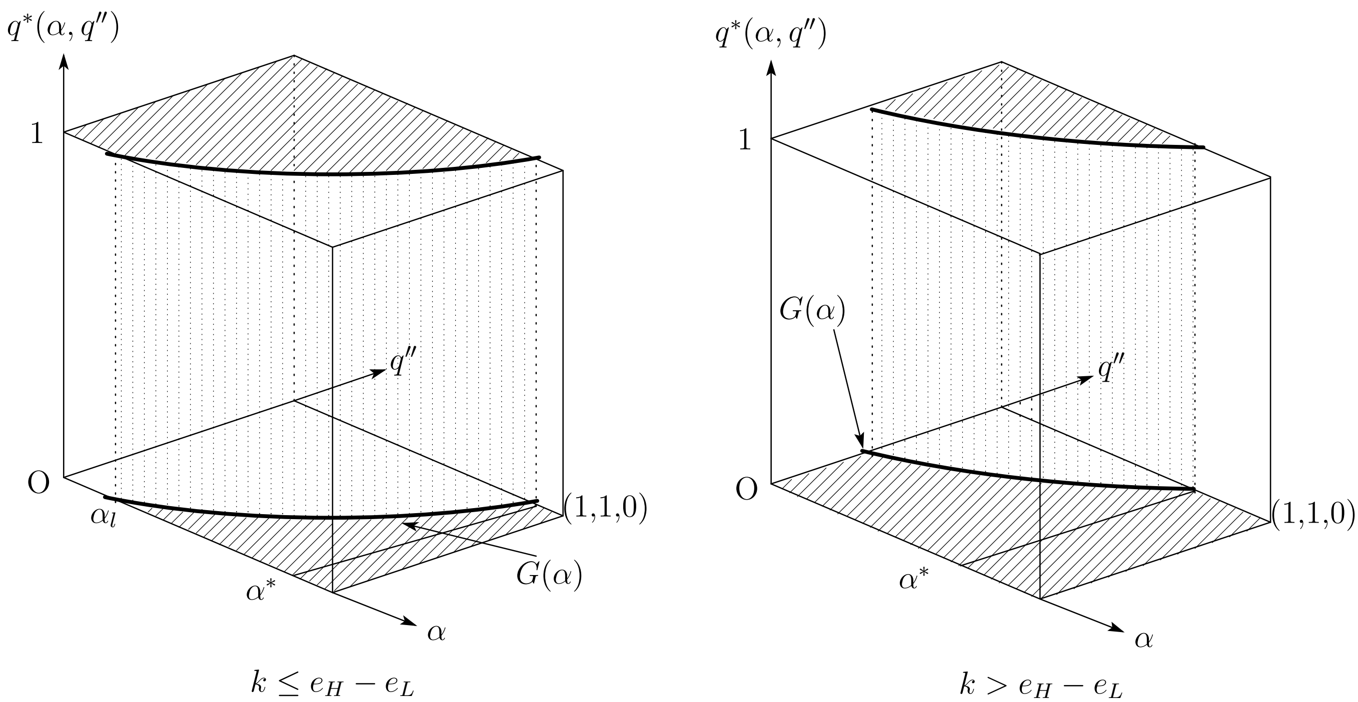

Figure 4: The best response correspondences $q^*(\alpha, q'')$ of player A.

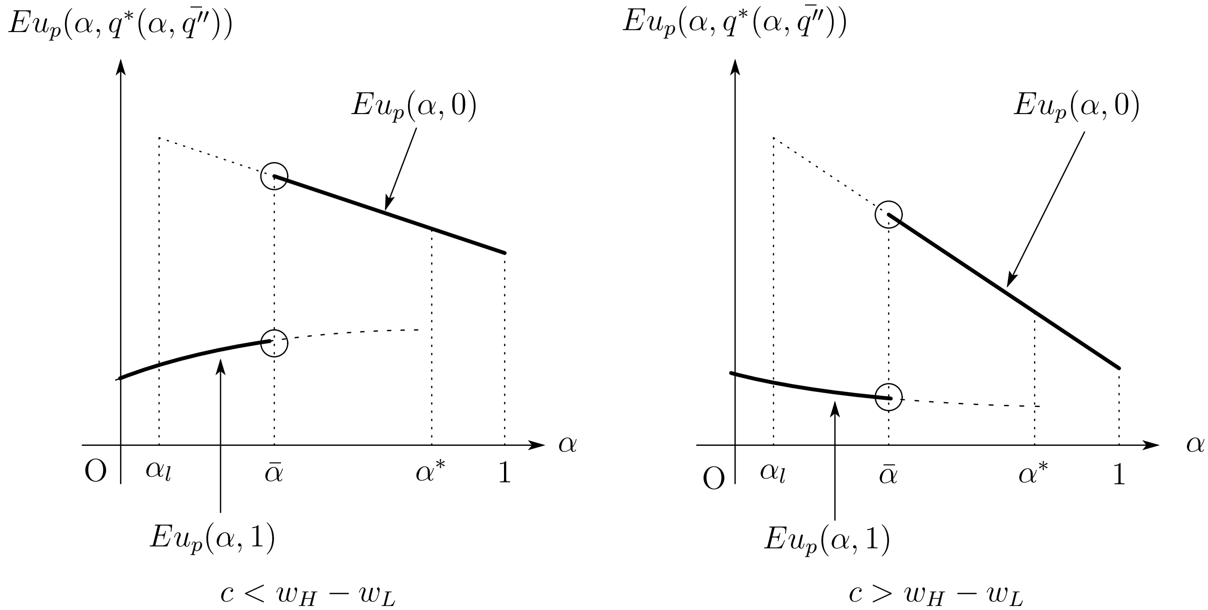

Figure 5: The expected payoff of player P for a fixed second order belief $\bar{q''} = G(\bar{\alpha})$.

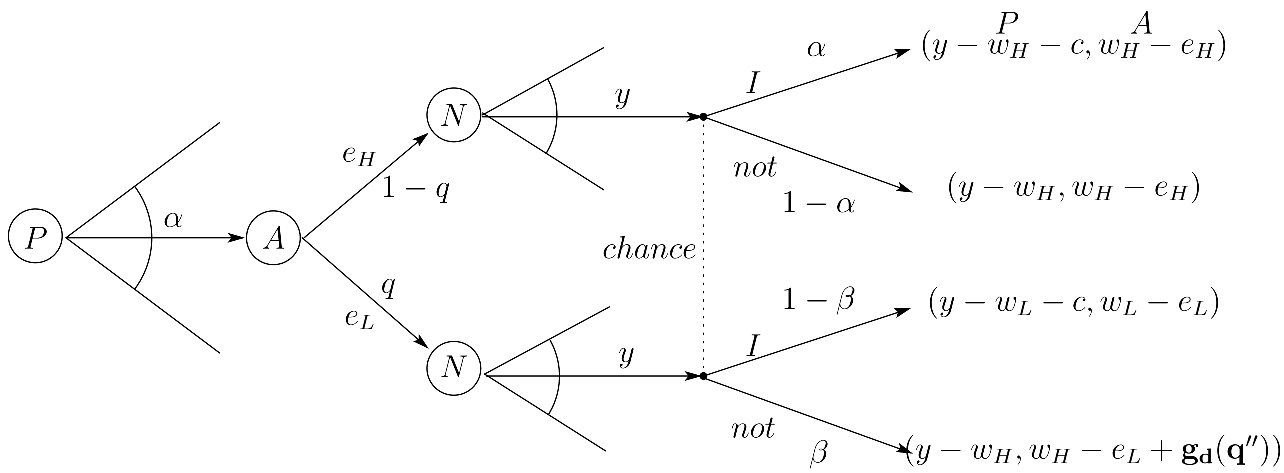

Figure 6: The model with an impulse to deceive.

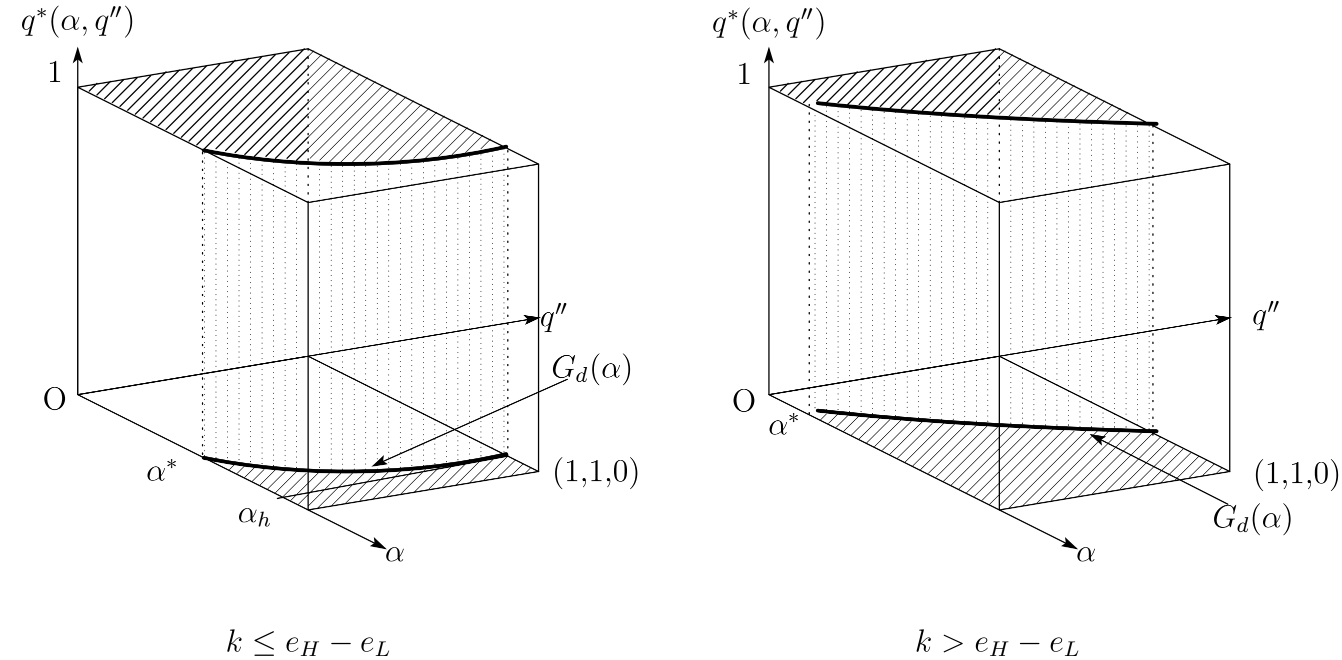

Figure 7: The best response correspondence $q^*(\alpha, q'')$ of player A.

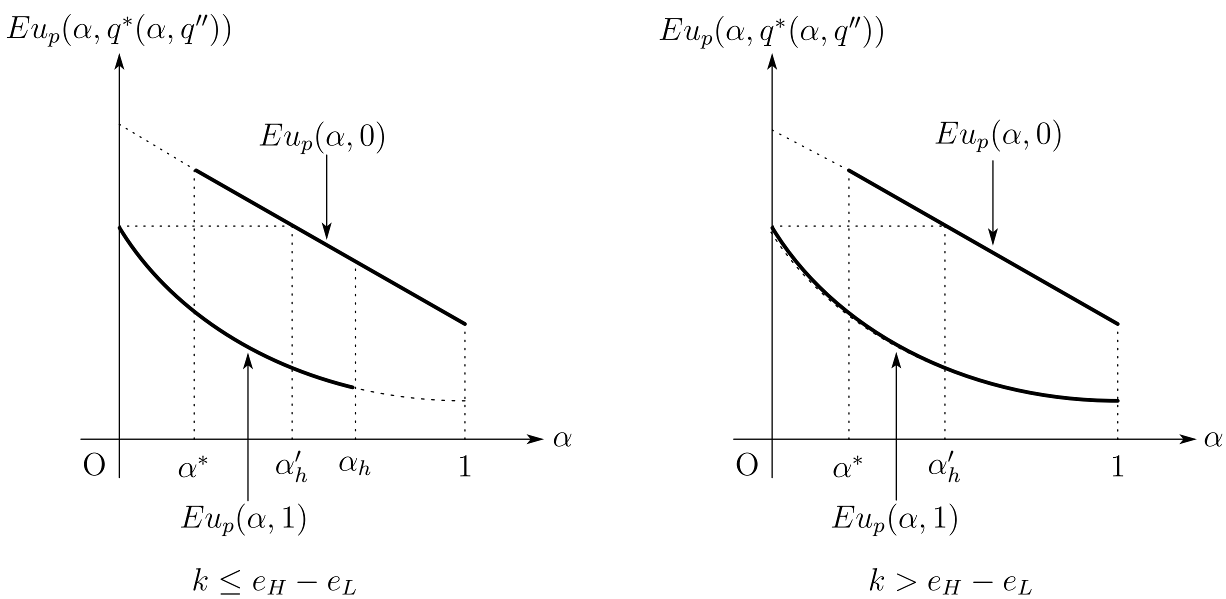

Figure 8: The possible expected payoffs of player P in case that there is $\alpha'_h \in (\alpha^*, \alpha_h)$ such that $Eu_p(\alpha'_h, 0) = Eu_p(0, 1)$.