Evolutionarily Stable Self-confidence Bias and Fixed Costs for Self-reproduction*

Masakazu Fukuzumi

1. Introduction

In the canonical models of Bayesian games each player with private information correctly knows his own type. If we recognize the implications of game theory as normative suggestions, the complete knowledge about his own type is reasonable assumption. On the other hand, if we turn to the empirical aspect of game theory, each player's knowledge about his own type would not necessarily be correct. Players' deduction about his own type might not necessarily be correct.

Möbius et al. (2012) is an experimental study and finds self-confidence bias of subjects about their own ability. Zhang (2013) considers a following evolutionary scenario in which a single agent decides his effort level for self-reproduction and justifies such bias successfully. Before the agent's decision, Nature assigns to the agent his type of ability for self-reproduction according to a probability distribution over the set of his own types. The set of his own types consists of high type and low type. The agent knows this probability distribution. Unlike the usual game-theoretic model, the agent can not observe directly this true type assigned by Nature. After Nature has picked up the true type, a noisy signal about his own type is generated through a mechanism that is a probability distribution conditional on that true type. The agent is assumed to be able to observe the signal and knows the conditional probability distribution. For example GPA scoring might be such a mechanism that generates the noisy signal about each student's ability. Given the signal, the agent forms a belief about his own ability.

With this belief, the agent decides the effort level for self-reproduction. In order to survive natural selection, the agent chooses the effort level so that his expected material payoff is maximized. This is as if Nature chooses the effort level that maximizes the expected fitness

* I would like to thank Yosuke Yoshioka (University of Tsukuba) for his valuable suggestions on proofs in this paper and Maho Suzuki for checking validity of results in this paper with numerical examples.

of a risk neutral agent based on a Bayesian belief about the agent's ability which is consistent with probability distributions generating the signals that the agent catches. However the real agent has his utility function that is not necessarily risk neutral. If the agent had a risk averse utility function, then the effort level chosen to maximize its expected utility might be different from the effort level that maximize the material payoff with Bayesian reasoning. In such a case the belief about the ability for self-reproduction should be adjusted to a non-Bayesian belief by the agent. Consequently, it is shown that the adjusted belief has self-confidence bias.

Consider, for example, a firm that is run by a risk-averse manager. If we follow the argument above, that manager is realizing the appropriate amount of investment to survive the competition by overestimating his firm's production technology. In the production process of such firms, there might be fixed costs. In our paper we introduce such a fixed cost for self-production into the model of Zhang (2013) and investigate its impact on the self-confidence bias.

Remaining part of this paper is organized as follows. Section 2 presents Zhang (2013) equipped with a fixed cost for self-reproduction. In Section 3 we show that depending types of risk averse, i.e., CRRA or CARA, the fixed cost has sharply different impacts on self-confidence bias. Section 4 is our concluding remarks.

2. Model

2.1 A system of generating signals

Let $T = \{H, L\}$ be the set of possible types of an agent. Each type represents his ability for self-production. At the beginning Nature picks up a type H (L) of the agent with probability μ_0 (resp.$1-\mu_0$). We suppose that the agent can not directly observe his own true type $t \in T$, but observe a signal $s \in T$. This signal is generated through publicly known conditional probabilities $p_1 = p(s = H|t = H)$ and $p_2 = p(s = H|t = L)$. Combining the knowledge about the prior μ_0, the conditional probabilities p_1, p_2 above, and the observed signal $s \in T$, the agent could form his belief μ that his true type is H. One of possible beliefs is the Bayesian posterior belief. Let μ^B denote the Bayesian posterior belief that his true type is H. Using logit$(\cdot)$, we can compactly write down all information about the Bayesian belief μ^B as follows:

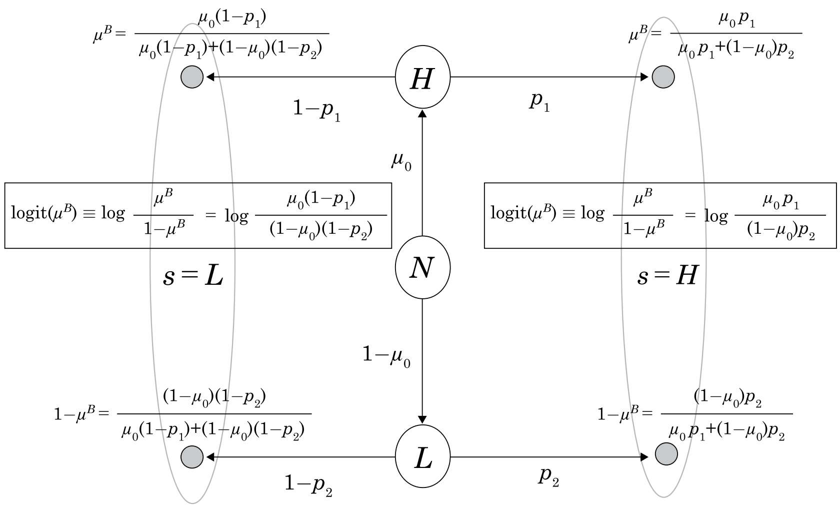
$$\text{logit}\left(\mu^B\right) = \text{logit}\left(\mu_0\right) + \mathbf{1}_{s=H}\lambda_H + \mathbf{1}_{s=L}\lambda_L,$$

where $\text{logit}(\mu) = \log(\frac{\mu}{1-\mu}), \lambda_H = \log(\frac{p_1}{p_2}), \lambda_L = \log(\frac{1-p_1}{1-p_2})$ and each $\mathbf{1}_{s\in T}$ is an indicator function of the signal $s \in T$ of which value is 1 or 0. Figure 1 illustrates this system generating the signals and the way to form the Bayesian posterior beliefs in the system.

2.2 Evolutionarily stable effort level

We consider a situation in which the agent with a belief about his own type chooses an

Figure 1 The system of generating signals and the derived Bayesian posterior beliefs



effort level to produce his own fitness i.e., material payoff. Let $a \in \mathbf{R}_{++}$ denote an effort level chosen by the agent. We assume a production functions for producing the material payoff $f(a,t) \in \mathbf{R}_{++}$ such that $f(a,H) > f(a,L)$ for each effort level $a \in \mathbf{R}_{++}$ and $f' > 0$, $f'' < 0$ (with $\lim_{a \to 0} f'(a,\cdot) = \infty$ and $\lim_{a \to \infty} f'(a,\cdot) = 0$.). Moreover we assume the cost function for producing the material payoff $C(a) \equiv c(a) + F$ in which $c(a) \in \mathbf{R}_{++}$ is the variable cost such that $c' > 0$, $c'' > 0$ and $F \in \mathrm{R}_{+}$ denotes the fixed cost. Nature selects the optimal level of action a^* maximizing his fitness $u_N(a)$ under the Bayesian belief μ^B so that

$$a^* \in \arg\max_a u_N(a) = \mu^B(f(a,H) - C(a)) + (1-\mu^B)(f(a,L) - C(a)).$$

Let f_a denote $\frac{\partial f}{\partial a}$. The first order condition for the Nature's optimization problem above becomes

$$\frac{\mu^B}{1-\mu^B} = \left| \frac{f_a(a^*,L) - c'(a^*)}{f_a(a^*,H) - c'(a^*)} \right| \tag{1}$$

where a^* is the optimal level of action under the Bayesian posterior belief μ^B.

2.3 Evolutionarily stable non-Bayesian belief μ^*

Whereas Nature maximizes the agent's fitness, we suppose that the agent maximizes his *expected utility* $u_A(a)$ with some belief μ. This belief μ is regarded as a subjective probability that the agent believes that his own type is H and it might not necessarily coincide with the Bayesian belief μ_B. We assume that the agent has a von Neumann-Morgenstern function $u : \mathbf{R} \to \mathbf{R}$ with $u'' < 0$, that is, the agent has risk averse preferences. The optimization problem for the agent is given by

$$\max_a u_A(a) = \mu u(f(a,H) - C(a)) + (1-\mu)u(f(a,L) - C(a)).$$

The first order condition of this problem implies

$$\frac{\mu}{1-\mu} = \left| \frac{f_a(a,L) - c'(a)}{f_a(a,H) - c'(a)} \right| \cdot \frac{u'(f(a,L) - c(a) - F)}{u'(f(a,H) - c(a) - F)}. \tag{2}$$

Given the optimal level of effort a^* determined implicitly in (1), the agent is assumed to adjust his belief μ to be consistent with the condition (2). Substituting the optimal level of effort a^* into the condition (2), we get the adjusted belief μ^* satisfying

$$\frac{\mu^*}{1-\mu^*} = \left| \frac{f_a(a^*,L) - c'(a^*)}{f_a(a^*,H) - c'(a^*)} \right| \cdot \frac{u'(f(a^*,L) - c(a^*) - F)}{u'(f(a^*,H) - c(a^*) - F)}. \tag{3}$$

From (2) and (3), we get the logit representation of difference between the adjusted belief μ^* and the Bayesian belief μ^B as follows.

$$\text{logit}(\mu^*) - \text{logit}(\mu^B) = \log\left[\frac{u'(f(a^*,L) - c(a^*) - F)}{u'(f(a^*,H) - c(a^*) - F)} \right].$$

Since $u'' < 0$ and $f(a^*,H) > f(a^*,L)$, we have $\frac{u'(f(a^*,L)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)} > 1$, that is, $\text{logit}(\mu^*) - \text{logit}(\mu^B) > 0$. We see that risk averse agents tend to have self-confidence bias (Zhang, 2013). The amount of $\log[\frac{u'(f(a^*,L)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)}]$ is called the *magnitude of self-confidence bias* for a fixed cost F and denoted by $B(F)$.

3. Analysis

The magnitude of self-confidence bias $B(F)$ is a function of the fixed cost F for self-reproduction. We show that depending on the type of risk aversion the impact of this fixed cost on the magnitude of self-confidence bias is shaply different. We consider two different classes of risk averse utility functions. One is the constant relative risk aversion (CRRA) utility function and the other is the constant absolute risk aversion (CARA) one. A von Neumann Morgenstern utility function $u(\cdot)$ is CRRA if its coefficient of relative risk aversion $-\frac{xu''(x)}{u'(x)}$ is constant. Let ρ be the coefficient of relative risk aversion for a CRRA utility function. A von Neumann Morgenstern utility function $u(\cdot)$ is CARA if its coefficient of absolute risk aversion $-\frac{u''(x)}{u'(x)}$ is constant. Let α be the coefficient of relative risk aversion for a CARA utility function.

Theorem 1 *If the agent's utility function is the CRRA, then the magnitude of self-confidence bias $B(F)$ is strictly increasing in the fixed cost F for self-reproduction.*

Proof. Define $b(F)$ to be a function $\frac{u'(f(a^*,L)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)}$ of F, that is the variable part of $B(F)$. From (1), a^* does not depend on F i.e., $\frac{\partial a^*}{\partial F} \equiv 0$. So the following calculation becomes simple.

$$\frac{\partial b(F)}{\partial F} = \frac{-u''(f(a^*,L) - c(a^*) - F)u'(f(a^*,H) - c(a^*) - F) + u''(f(a^*,H) - c(a^*) - F)u'(f(a^*,L) - c(a^*) - F)}{(u'(f(a^*,H) - c(a^*) - F))^2}$$

$$= \frac{u''(f(a^*,L)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)} + \frac{u''(f(a^*,H)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)} \cdot \frac{u'(f(a^*,L)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)}$$

$$= \frac{u''(f(a^*,L)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)} \left\{ -1 + \frac{u''(f(a^*,H)-c(a^*)-F)}{u''(f(a^*,L)-c(a^*)-F)} \cdot \frac{u'(f(a^*,L)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)} \right\}.$$

Since $u' > 0$ and $u'' < 0$, $\frac{u''(f(a^*,L)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)} < 0$.

Our remaining task for for determining the sign of $\frac{\partial b}{\partial F}$ is to check the sign of the second half of the above formula (4). By arranging the second half of the above formula, we get

$$-1 + \frac{u''(f(a^*,H)-c(a^*)-F)}{u''(f(a^*,L)-c(a^*)-F)} \cdot \frac{u'(f(a^*,L)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)}$$

$$= -1 + \left[\frac{u'(f(a^*,L)-c(a^*)-F)}{u''(f(a^*,L)-c(a^*)-F)} \cdot \frac{1}{f(a^*,L)-c(a^*)-F} \right]$$

$$\cdot \left[\frac{u''(f(a^*,H)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)} \cdot (f(a^*,H)-c(a^*)-F) \right] \cdot \frac{f(a^*,L)-c(a^*)-F}{f(a^*,H)-c(a^*)-F}. \tag{4}$$

Both ingredients of $\frac{u'(f(a^*,L)-c(a^*)-F)}{u''(f(a^*,L)-c(a^*)-F)} \cdot \frac{1}{f(a^*,L)-c(a^*)-F}$ and $\left[\frac{u''(f(a^*,H)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)} \cdot (f(a^*,H)-c(a^*)-F)\right]$ of the second half of (4) are the coefficient of relative risk aversion $-\frac{xu''(x)}{u'(x)}$ of this agent's utility function. Since u is a CRRA utility function, the value of these coefficients is equal to some constant ρ. Substituting ρ into the formula (4), we get a formula $-1 + \frac{1}{\rho} \cdot \rho \cdot \frac{f(a^*,L)-c(a^*)-F}{f(a^*,H)-c(a^*)-F}$. Since we have assumed that $f(a,H) > f(a,L)$ for each effort level $a \in \mathbf{R}_{++}$, $\frac{1}{\rho} \cdot \rho \cdot \frac{f(a^*,L)-c(a^*)-F}{f(a^*,H)-c(a^*)-F} < 1$, namely $-1 + \frac{1}{\rho} \cdot \rho \cdot \frac{f(a^*,L)-c(a^*)-F}{f(a^*,H)-c(a^*)-F} < 0$.

From $\frac{u''(f(a^*,L)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)} < 0$ and $-1 + \frac{u''(f(a^*,H)-c(a^*)-F)}{u''(f(a^*,L)-c(a^*)-F)} \cdot \frac{u'(f(a^*,L)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)} < 0$, we get $\frac{\partial b(F)}{\partial F} > 0$ for each F. ■

Theorem 2 *If the agent's utility function is CARA, then the magnitude of self-confidence bias $B(F)$ is irrelevant with the level of fixed cost F for self-reproduction.*

Proof. Note that $B(F) = \log[\frac{u'(f(a^*,L)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)}]$ Define $b(F)$ to be a function $\frac{u'(f(a^*,L)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)}$ of F. We check the sign of $\frac{\partial b}{\partial F}$. From (1), a^* does not depend on F i.e., $\frac{\partial a^*}{\partial F} \equiv 0$. So the following calculation becomes simple.

$$\frac{\partial b(F)}{\partial F} = \frac{u''(f(a^*,L)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)} \left\{ -1 + \frac{u''(f(a^*,H)-c(a^*)-F)}{u''(f(a^*,L)-c(a^*)-F)} \cdot \frac{u'(f(a^*,L)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)} \right\}.$$

To check the sign of $\frac{\partial b(F)}{\partial F}$, we focus on the sign of the second half of the above formula. By arranging this part, we have

$$-1 + \frac{u''(f(a^*,H)-c(a^*)-F)}{u''(f(a^*,L)-c(a^*)-F)} \cdot \frac{u'(f(a^*,L)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)}$$

$$= -1 + \left[\frac{u''(f(a^*,H)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)} \right] \cdot \left[\frac{u'(f(a^*,L)-c(a^*)-F)}{u''(f(a^*,L)-c(a^*)-F)} \right].$$

Since the agent's utility function is CARA, let α be the constant value of the coefficient of absolute risk aversion $-\frac{u''(x)}{u'(x)}$. Substituting α to that arranged formula, we get $-1 + \alpha \cdot \frac{1}{\alpha} = 0$. That is, $\frac{\partial b(F)}{\partial F} = 0$. ■

4. Concluding remarks

When the agent is constant relative risk averse, the level of the fixed cost for self production has sharp impact on the bias of his self-confidence about his own ability. On the other hand, the level of the fixed cost has no effect on the bias for the agent who is constant absolute risk averse. Both of these facts above are deduced from the same evolutionary model of Zhang (2013). Therefore we would be able to test the validity of this evolutionary model of the bias with adequate designed experiments.

Regardless from whether the agent is risk averse or not, the effort level chosen by agent is assumed to be the optimal level such that his material payoff is maximized. The agent modifies his belief from Bayesian belief to attain this optimal effort level. This scenario is similar to the approach of the preference evolution (Samuelson, 2001). In the model of preference evolution the preference is adjusted but the belief is not. In our model the belief is adjusted but the preference is not. Which of the preferences and beliefs is more strongly subject to evolutionary pressure? This is one of most interesting topics for future research.

References

Möbius, M. M., Niederle, M., Niehaus, P., Rosenblat, T. S. (2012) “Managing self-confidence: Theory and experimental evidence," NBER Working Paper, No.17014.

Samuelson L. (2001) “Introduction to the evolution of preferences," *Journal of Economic Theory*, **97**, 225-230.

Zhang, H. (2013) “Evolutionary justifications for non-Bayesian beliefs," *Economics Letters*, **121**, 198-201.

(Associate Professor, Faculty of Humanities and Social Sciences,
University of Tsukuba, Dr. of Economics)