On the characteristics of evolutionarily stable self-confidence bias*

Yosuke Yoshioka[†] Masakazu Fukuzumi[‡]

Abstract

We investigate the characteristics of evolutionarily stable self-confidence bias that is non-Bayesian belief of a single decision maker. It is known that whenever a decision maker is risk averse, he has the self-confidence bias for his own ability of his self-reproduction. We show that the size of the bias is invariant in affine transformations in the utility function, which indicates that the bias has a similar property to expected utility hypothesis. Moreover, introducing fixed costs into self-reproduction process, we investigate the effects of time horizontal perspectives on the bias. We find that the size of the bias of a decision maker with CRRA utility function tends to increase in the short run. In contrast, the fixed cost for self-reproduction is irrelevant with the size of the bias for another type decision maker with CARA utility function.

JEL classification: C73, D83
Keywords: non-Bayesian belief, CRRA, CARA, fixed cost, affine transformation

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 that 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. A 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 over the set of types. For example GPA scoring might be such a mechanism that generates

*This work was supported by JSPS KAKENHI Grant Number JP19K01540.
[†]Graduate School of Humanities and Social Sciences, University of Tsukuba
[‡]Faculty of Humanities and Social Sciences, University of Tsukuba

a 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 of a risk neutral agent based on a Bayesian belief about the agent's ability which is consistent with the two probability distributions above generating the signals. 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 the Bayesian belief. In such a case the belief about the ability for self-reproduction should be adjusted to a non-Bayesian belief by the agent himself to attain the optimal effort chosen by Nature.

We, first, examine the effect of affine transformations in the utility function on the self-confidence bias because the transformation is one of the key concepts for the expected utility hypothesis. Second, we turn to investigate the effects of time horizontal perspectives on the bias and its applicability of our argument to standard economic theory. The self-reproduction's cost might depend on the time horizon under consideration, that is, the short-run or the long run. In the short-run the agent might bears the fixed costs for his self-reproduction. We introduce the fixed cost for self-reproduction into the model of Zhang (2013) and investigate its impact on the self-confidence bias. [1]

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 affine transformations has no effect on the size of the bias and depending on 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 a 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;

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

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

[1] Suzuki (2020) points out that this introduction of fixed costs increases the self-confidence bias in some case with a numerical example. Our paper summarizes some of the earlier study of Fukuzumi (2020) , one of our paper's authors, that analyzes the relationship between fixed costs and the size of the bias, along with the other properties of the bias found in our paper.

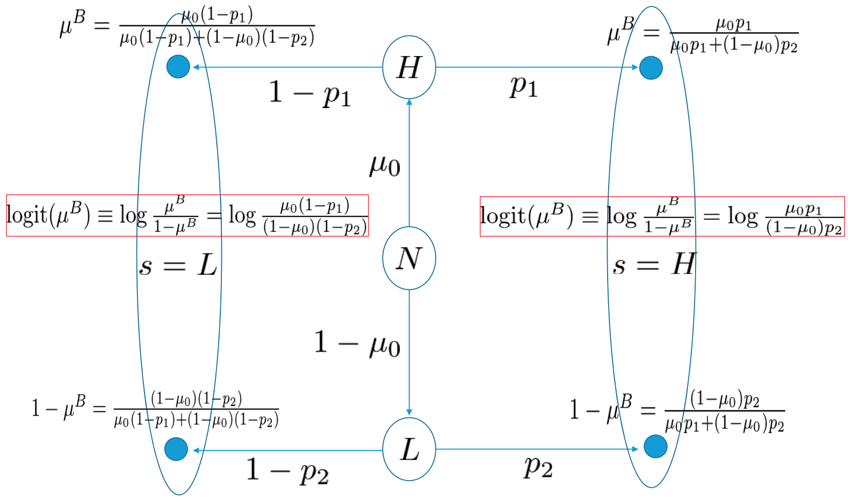

Figure 1: The system of generating signals and the Bayesian posterior beliefs derived from it.

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 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 $f' > 0$, $f'' < 0$. 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.

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

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

3 Analysis

In expected utility hypothesis, affine transformations to utility functions have no effect on decision makers' behavior. We investigate whether or not the transformation has any effect on the self-confidence bias. Let $U(\alpha,\beta;F)$ be the expected value of an affine transformed utility function $\alpha u(\cdot) + \beta$ ($\alpha, \beta \in \mathbf{R}, \alpha > 0$) with a belief μ. The agent's optimal behavior solves

$$\max_a U(\alpha,\beta;F) = \mu\{\alpha u(f(a,H) - C(a)) + \beta\} + (1-\mu)\{\alpha u(f(a,L) - C(a)) + \beta\}.$$

The first order condition of this problem implies

$$\begin{aligned}\frac{\mu}{1-\mu} &= \left|\frac{f_a(a,L) - c'(a)}{f_a(a,H) - c'(a)}\right| \cdot \frac{\alpha u'(f(a,L) - c(a) - F)}{\alpha u'(f(a,H) - c(a) - F)} \\ &= \left|\frac{f_a(a,L) - c'(a)}{f_a(a,H) - c'(a)}\right| \cdot \frac{u'(f(a,L) - c(a) - F)}{u'(f(a,H) - c(a) - F)}. \qquad (2)\end{aligned}$$

Note that this result is the same as that in Section 2.3. Substitute the optimal action a^* under the Bayesian posterior belief μ^B, we get the non-Bayesian belief μ^* which is also the sames as that in Section 2.3. Thus the affine transformation of utility function does not affect the size of the self-cofidence bias $B(F)$. In the following we analyze the size of bias without taking affine tranformations into account.

The size 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 size of self-confidence bias is shaply different. We consider two typical 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.

Observation 1. For a CRRA function $u(x) = x^{1-\rho}/(1-\rho), \rho \geq 1$, its size of self-confidence bias $B(F)$ is an increasing function of F.
In this case $B(F) = \rho \log\left|\frac{f(a^*,H)-c(a^*)-F}{f(a^*,L)-c(a^*)-F}\right|$. Let $b(F) \equiv \frac{f(a^*,H)-c(a^*)-F}{f(a^*,L)-c(a^*)-F}$. Since $\frac{\partial b(F)}{\partial F} = \frac{-(f(a,L)-c(a)-F)+(f(a,H)-c(a^*)-F)}{(f(a^*,L)-c(a^*)-F)^2} > 0$, we get this fact. $\square$
We show that this finding holds more generally as follows.

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

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

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

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

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

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 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

$$-1 + \frac{u''(f(a^*,H)-c(a^*)-F)}{u''(f(a^*,L)-c(a^*)-F)} \cdot \frac{u'(f(a^*,L)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)}$$
$$= -1 + \left[\frac{u'(f(a^*,L)-c(a^*)-F)}{u''(f(a^*,L)-c(a^*)-F)} \cdot \frac{1}{f(a^*,L)-c(a^*)-F}\right] \cdot \left[\frac{u''(f(a^*,H)-c(a^*)-F)}{u'(f(a^*,H)-c(a^*)-F)} \cdot\right.$$
$$\left.(f(a^*,H)-c(a^*)-F)\right] \cdot \frac{f(a^*,L)-c(a^*)-F}{f(a^*,H)-c(a^*)-F}. \tag{4}$$

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. Let ρ be the constant value of the coefficient of absolute risk aversion $-\frac{u''(x)}{u'(x)}$. 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. $\blacksquare$

Observation 2. For a CARA function $u(x) = K - \exp(-\tau x), \tau \geq 0$, its size of self-confidence baias $B(F)$ is irrelevant with F.
In this case $B(F) = \alpha[f(a^*,H) - f(a^*,L)]$. $\square$
This finding of irrelevance holds more generally as follows.

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

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

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

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 τ be the constant value of the coefficient of absolute risk aversion $-\frac{u''(x)}{u'(x)}$. Substituting τ to that arranged formula, we get $-1+\tau\cdot\frac{1}{\tau} = 0$. That is, $\frac{\partial b(F)}{\partial F} = 0$. ■

4 Concluding remarks

We have shown that the evolutionarily stable self-confidence bias is invariant in the affine transformation. This fact indicates that this non-Bayesian decision theory is partly similar to expected utility hypothesis. Furthermore, we have found the fact that whenever a CRRA decision maker's bias tends to increase in the short-run. This fact might corresponds to daily feelings in which bounded rational behaviors are likely to be seen in the short-run.

The agent modifies his belief from Bayesian one to attain the optimal effort level chosen by Nature. This scenario is partly similar to the approach of the preference evolution models in which preferences are adjusted to maximize the material payoff but the beliefs are not (Samuelson, 2001). Which of the preferences and beliefs are more strongly subject to evolutionary pressure? This might be one of most interesting topics for future research.

References

Fukuzumi, M (2021) "Evolutionarily stable self-confidence bias and fixed costs for self-reproduction," *The Journal of Economics*, **61**, Chuo University, forthcoming.

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.

Suzuki, M. (2020) "Hi Bayesian sinnenn ・ ninnchi-baiasu no seisei ni tuite no shinnkateki rironn bunnseki," Unpublished undergraduate dissertation (in Japanese), University of Tsukuba.

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