한국경제학보 제27권 제1호
The Korean Journal of Economics Vol. 27, No. 1 (Spring 2020)

Signaling, Prior-bias and Welfare Implications*

Eunjee Joo** · Yong-Ju Lee***

Abstract

Although most economics and social science textbooks describe Bayes' rule as the basic principle of probabilistic inference combining prior information with additional information from new observations, it is widely observed in the laboratory experiments that people's information processing systemically deviates from Bayes' rule. Here we illustrate how receiver's prior-biasedness affects players' strategic incentives and welfare implications in the "costless" and "costly" models of signaling. From these examples we draw an interesting welfare implication that the receiver's prior-bias may work as a welfare-enhancing device even in the costly signaling environments. That is, contrary to the widespread beliefs that non-Bayesian inference is a sign of human irrationality and a social cost, we show that a behavioral bias can be beneficial.

KRF Classification : B030200, B030904
Keywords : Non-Bayesian updating, Prior-bias, Signaling, Beer-Quiche game, Job market signaling

---

* This work was supported by the 2019 Yeungnam University Research Grant.

** First Author, Professor, Department of Youth Culture & Counseling, Hoseo University, 12 Hoseodae-gil, Cheonan 31066, Korea, e-mail: jooej@hoseo.edu

*** Corresponding Author, Associate Professor, School of Economics and Finance, Yeungnam University, 280 Daehak-ro, Gyeungsan 38541, Korea, e-mail: yongjulee@ynu.ac.kr

Ⅰ. Introduction

Most economics and social science textbooks describe Bayes' rule as the basic principle of statistical inference combining prior information with additional information that comes from new observations, and consider non-Bayesian updating to be an irrational bias which leads to sub-optimal outcome. In contrast, laboratory evidences show that people's information processing systemically deviates from Bayes' rule (e.g., Benjamin, 2019; Charness and Dave, 2017; Charness and Levin, 2005). Nonetheless, the theoretic research on non-Bayesian updating in economics has been mainly confined to decision-theoretic environments (e.g., Epstein, 2006; Epstein et al., 2008, Benjamin, 2019). Hence, there still remain the questions of how the non-Bayesian bias works in terms of players' strategic incentives and welfare implications in other strategic settings.

Recently we have witnessed that the research on the effects of non-Bayesian biases in the strategic environments began emerging. Lee, Lim and Zhao (2020) examine how non-Bayesian updating, specifically prior-biased inference, affects players' strategic incentives and welfare implications in the model of strategic information transmission. The introduction of prior-bias creates the conflicting tradeoff. On the one hand, the decisions made by a receiver are distorted by her prior-bias. On the other hand, knowing the receiver's biasedness, a sender has an incentive to deliver more informative signals. Thus, whether prior-bias improves social welfare depends on the relative size of these two effects. They find that prior-bias is a welfare-enhancing device and the social welfare in the model with prior-bias exceeds that in the standard model without prior-bias. Their findings are also related to Heifetz, Shannon and Spiegel (2007)'s result that, albeit not about non-Bayesian updating, in almost every game and for almost every type of distortion of a player's actual payoffs, some degree of distortion is beneficial to the player because of the resulting effect on opponents' play.

Another paper, de Clippel and Zhang (2019) expand the analysis of information design (also referred to as Bayesian persuasion) by accommodating non-Bayesian receiver who is prone to mistakes in probabilistic inference. They find that even if the receiver is non-Bayesian, the sender can effectively persuade her to take a desired action by controlling her informational environment.

In this paper, applying Lee, Lim and Zhao (2020)'s idea, we investigate how the presence of prior-bias changes players' incentives in the models of signaling and check what kind of implication we can draw from those. The positive prior-bias, as a simplified version of confirmation bias, captures agent's biased inferences drawn in favor of the current belief. This prior bias in belief updating has long been documented in the psychology and experimental literature (Pitz, Downing, and Reinhold, 1967; Geller and Pitz, 1968; Pitz, 1969, etc).

Specifically, we illustrate the notion of prior-biased inference and welfare implications in the restricted cheap talk game and two examples costly signaling models, the beer-quiche game (Cho and Kreps, 1987) and the Spence (1973)'s job market signaling. We show that in certain conditions non-Bayesian bias may lead to the emergence of new equilibria and better payoffs or even Pareto improvements.

The remainder of the paper is organized as follows. Section 2 introduces the concept of prior-bias and the equilibrium concept in the presence of prior-bias. In sections 3, we apply the idea of prior-biased inference to the binary cheap talk game, the Beer-Quiche game and Spence's job market signaling environment. We conclude in section 4.

Ⅱ. Prior-bias and Related Equilibrium

We assume that the receiver is prior-biased as in Epstein (2006) that provides an axiomatic foundation and representation theorem of prior-bias.[1)] Following Epstein (2006), the receiver with prior-bias draw inferences in the following way: for a signal m, ·

$$\psi(\cdot\,|m)=(1-a)\mu(\cdot\,|m)+ap_0(\cdot) \tag{1}$$

where $p_0(\cdot)$ is the (common) prior and $\mu(\cdot\,|m)$ is the Bayesian update of $p_0(\cdot)$. $a\in(0,1)$ indicates the degree of prior-bias.

The receiver draws inferences in favor of current belief and the sender knows about this. We abstract from the consideration of higher order beliefs and assume the degree of prior-bias a is common knowledge. So, the sender takes the prior-bias of the receiver into account in his signaling decision and the sophisticated prior-biased receiver knows that she cannot commit to fully updating her beliefs based on Bayes' rule.[2)]

Following Lee, Lim and Zhao (2020), we apply an equilibrium concept to signaling games, called *perfect prior-biased equilibrium*:[3)]

1. The sender optimizes against the receiver's strategy.

---

1) Epstein (2006) introduces a more general concept of prior-bias. First, in Epstein (2006), the parameter $a(m)$ can vary over a signal m. Second, $a(m)$ can be negative: Negative prior-bias. Here we only focus on the positive prior-bias (or underreaction to new information) with the constant parameter value. And for the ease of exposition and to avoid arbitrariness, we assume the parameter a constant. We thank an anonymous referee for pointing out this.

2) Whether the receiver is sophisticated or not does not matter for equilibrium characterization as long as the sender is aware of the prior-bias of the receiver. We assume the receiver's sophistication to avoid the controversy on how expected welfare should be evaluated.

3) Eyster and Rabin (2005) introduces the concept of x-cursed equilibrium, which is equivalent to the perfect prior-biased equilibrium in the signaling game environment. We discuss this in the Appendix.

2. The receiver draws inferences based on the posterior belief $\psi(\bullet|m)$ whenever possible.
3. The receiver's strategy constitute a Nash equilibrium, given $\psi(\bullet|m)\forall m$.

The only difference from perfect Bayesian equilibrium is that the Bayesian posterior belief $\mu(\bullet|m)$ that the receiver has whenever possible is replaced by prior-biased posterior $\psi(\bullet|m)$.[4)]

Ⅲ. Application to the Signaling Games

In some strategic environments, information updating plays a crucial role, such as signaling and learning. Here we confine our attention to the models of signaling which share similar communication environment.

1. Restricted Cheap Talk

In this section, unlike Lee, Lim and Zhao (2020), we consider a restricted version of cheap talk where the players' choice is limited. The model considered in this subsection is different from a proper cheap talk game in that the number of elements in the message space is not greater than the number elements of the power set of the state space. So the main goal of this subsection is confined to illustrate the characterization of perfect prior-biased equilibrium.

There are two players, a sender (S) and a receiver (R), in the cheap talk setting. The sender is privately informed about the state of nature $\theta\in\{0,1\}$. The sender sends a message $m\in M=\{0,1\}$ to the receiver who then takes an action $y\in Y=[0,1]$. The common prior is that each state happens with equal probability $1/2$. The payoffs are

4) Refer to Lee, Lim and Zhao (2019) for a formal definition of the equilibrium.

as follows: $U_R(\theta,y)=-(\theta-y)^2$ and $U_S(\theta,y,b)=-(\theta+b-y)^2$ where $b>0$ is the parameter measuring the conflict of interest between the two players. We assume that the receiver is prior-biased as explained above.

The sender's strategy $\sigma:\{0,1\}\rightarrow\Delta M$ is a type-dependent message plan that maps his type $\theta\in\{0,1\}$ to a message $m\in M$. The receiver's strategy $y:M\rightarrow Y$ is a message-contingent action plan that maps a message $m\in M$ she could receive to an action $y\in Y$. If the sender chooses m according to his strategy $\sigma(m|\theta)$, then after receiving message, the prior-biased receiver's posterior belief about state θ is

$$\psi(\theta|m)=(1-a)\frac{\frac{1}{2}\sigma(m|\theta)}{\frac{1}{2}(\sigma(m|\theta)+\sigma(m|1-\theta))}+\frac{1}{2}a.$$

A pair of strategies, $\{\sigma,y\}$ together with a belief system ψ constitutes a perfect prior-biased equilibrium.

Here we characterize all the equilibria of this game. Depending on the values of a and b, there exist different types of equilibria.

1. Separating equilibrium can emerge if $b<\frac{1}{2}$:

In the equilibrium, $\sigma(m|\theta)=1$ if $m=\theta$, $\sigma(m|\theta)=0$ otherwise; $y(0)=\frac{1}{2}a$, $y(1)=1-\frac{1}{2}a$; and the (prior-biased) posterior beliefs supporting the receiver's choice are

$$\begin{cases}\psi(0|0)=(1-a)\bullet 1+\frac{1}{2}a=1-\frac{1}{2}a\\ \psi(1|0)=(1-a)\bullet 0+\frac{1}{2}a=\frac{1}{2}a\\ \psi(0|1)=(1-a)\bullet 0+\frac{1}{2}a=\frac{1}{2}a\\ \psi(1|1)=(1-a)\bullet 1+\frac{1}{2}a=1-\frac{1}{2}a.\end{cases}$$

2. Partially separating equilibrium can emerge if $\frac{1-a}{4} < b < \frac{1}{2}$:

In the equilibrium, $\sigma(1|1)=1$, $\sigma(1|0)=\frac{1-2b}{2b-a}$; $y(0)=\frac{1}{2}a$, $y(1)=2b-\frac{1}{2}a$; and the posterior beliefs supporting the receiver's choice are

$$\begin{cases} \psi(0|0)=(1-a)\cdot 1+\frac{1}{2}a=1-\frac{1}{2}a \\ \psi(1|0)=(1-a)\cdot \frac{\sigma(1|0)}{1+\sigma(1|0)}+\frac{1}{2}a=1-2b+\frac{1}{2}a \\ \psi(0|1)=(1-a)\cdot 0+\frac{1}{2}a=\frac{1}{2}a \\ \psi(1|1)=(1-a)\cdot \frac{1}{1+\sigma(1|0)}+\frac{1}{2}a=2b-\frac{1}{2}a. \end{cases}$$

3. There always exists babbling equilibrium regardless of the values of a and b.

By plugging equilibrium strategies, we can easily calculate the expected utilities of players.

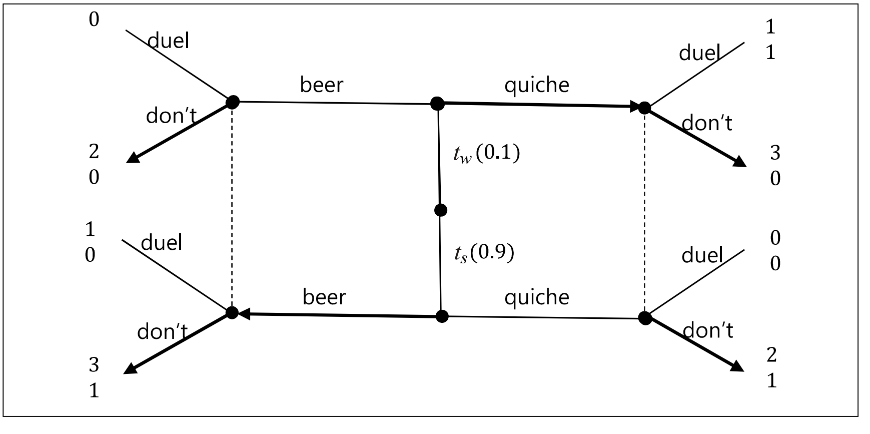
The expected utilities of players in the separating equilibrium deteriorates when introducing the receiver's biasedness to the model: $EU=-b^2-\frac{1}{4}a^2$ for the sender and $EU=-\frac{1}{4}a^2$ for the receiver in the perfect prior-biased equilibrium are lower than $EU=-b^2$ for the sender and $EU=0$ for the receiver in the perfect Bayesian equilibrium without prior-bias. The welfare result of this example is different from Lee, Lim and Zhao (2020) in that the social welfare achieved by the most informative equilibrium of the model with prior-bias exceeds the upper bound of social welfare characterized in the standard environment without prior-bias. The next two examples consider "costly" signaling environments: The beer-quiche game and the job market signaling game.

2. The beer-quiche game with prior-bias

The structure of beer-quiche game (Cho and Kreps, 1987) is similar to the example of cheap talk explained above. The signal m is irrelevant to the payoff functions in the cheap talk game, but it is critical in the signaling games. This example will clearly show how the prior-biasedness of receiver changes equilibrium outcomes and that in certain conditions behavioral biases may lead to a better payoff for the sender or even Pareto improvements.

The beer-quiche game is a simple two-player game with incomplete information about a sender's type. The sender is a wimp with probability 0.1 or is surly with probability 0.9 by nature's selection. He knows his own type and is faced with the choice of what breakfast to have between quiche and beer. If he is a wimp, he derives payoff 1 from quiche and 0 from beer. If he is surly, he gets payoff 1 from beer and 0 from quiche. After breakfast, the sender meets with the receiver who chooses whether to duel knowing what the sender had for breakfast, but not knowing the sender's type. The sender, regardless of his type, wishes that the receiver chooses not to duel: the sender gets incremental payoff 2 if the receiver chooses not to duel and 0 if the receiver duels. However, the receiver wishes to

〈Figure 1〉 The Beer-Quiche game



duel with the sender and get payoff 1 if and only if the sender is wimp. The game ends with the receiver's choice. By choosing breakfast effectively, the sender may deter the receiver from dueling.

As is well known in the literature, there are just two perfect Bayesian equilibrium outcomes in the game: either both types have beer for breakfast or both types have quiche, and the receiver don't duel. There is no separating equilibrium in the perfect Bayesian equilibria.

What happens if the receiver is prior-biased? First, note that there is no change in the pooling equilibrium. Both the types send the same message and no additional information is conveyed to the receiver, and thus the receiver's posterior beliefs on the equilibrium path is the same as his priors. Hence, every perfect Bayesian pooling equilibrium is perfect prior-biased pooling equilibrium.

Now we show that there emerge new perfect prior-biased *separating* equilibria depending on the value of a. The equilibrium actions are marked with thick arrows in Figure 1. Suppose that $a > 1/1.8$. Then, in the new (prior-biased) separating equilibrium, the wimp type orders quiche and the surly type orders beer, and the receiver, regardless of message, chooses to avoid the duel. The (prior-biased) posterior beliefs supporting the receiver's choice are

$$\begin{cases} \psi(\text{wimp}|\text{quiche}) = (1-a) \cdot 1 + a(0.1) = 1 - 0.9a \\ \psi(\text{surly}|\text{quiche}) = (1-a) \cdot 0 + a(0.9) = 0.9a \\ \psi(\text{wimp}|\text{beer}) = (1-a) \cdot 0 + a(0.1) = 0.1a \\ \psi(\text{surly}|\text{beer}) = (1-a) \cdot 1 + a(0.9) = 1 - 0.1a. \end{cases}$$

In the above example, the prior belief of the receiver says that the sender is very likely to be surly (that is, $p_0(s) = 0.9$). With a presence of the high degree of prior-bias (here $a > 1/1.8$), even observing quiche for breakfast, the receiver still maintains the belief that the sender is likely to be surly and chooses not to duel. Then each type chooses her preferred breakfast. The key feature is that even if only the wimp type chooses quiche for breakfast in the equilibrium, ψ(surly | quiche) is not zero in the *mind* of the receiver.

The outcome described above is never a perfect Bayesian equilibrium outcome. Furthermore, this equilibrium outcome is the best attainable outcome for the sender, without making the receiver worse off, and suggests the potential welfare effects of prior-bias in the costly signaling environment as is the case in the costless cheap talk environment of Lee, Lim and Zhao (2020). We can find the similar result even after changing some (or all) of the payoffs or with some undetermined variables in the Beer-quiche game while keeping the main feature of the game as long as the probability of being surly and the value of a are high enough. What is important is to make the receiver maintain the belief that the sender is very likely to be surly.

Thus we have the following summary of results in the beer-quiche game: with the presence of prior-bias, (1) there may emerge perfect prior-biased *separating* equilibria with a high degree of prior-bias, (2) the new perfect prior-biased *separating* equilibrium gives the best attainable outcome for the sender, without making the receiver worse off.

3. Employer's Prior-bias and the Job Market Signaling

In this section we examine the welfare effects of prior-bias in the labor market signaling model due to Spence (1973), but here, unlike the original model, we consider the problem of education as a signal in a bilateral contracting setting with one prior-biased employer and one worker.

There are two types for the worker, $\theta_H > \theta_L > 0$. A worker of type θ produces output which is worth θ to the employer. There is a commonly known prior probability that the worker's productivity is high:

$$\mathrm{Prob}(\theta=\theta_H)=\lambda\in(0,1).$$

The worker can acquire education, $e\geq 0$, which is perfectly observable. Assume that productivity is assumed to be innate so that education does not affect productivity. In this sense, education is purely "wasteful". The cost of education $c(e,\theta)$ is type specific. For a signaling role of education, assumptions needed for the cost function are $c(0,\theta)=0$, $c_e(e,\theta)>0$, $c_{ee}(e,\theta)\geq 0$, $c_\theta(e,\theta)<0$, and $c_{e\theta}(e,\theta)<0$. For the low type it is costly mimic high education choices of the high type, which allow the high type to differentiate herself and obtain a higher wage. The worker's utility is a function of her wage, w, her education, e, and her productivity type θ:

$$u(w,e,\theta)=w-c(e,\theta).$$

Notice that different types face different education-wage tradeoffs by the single-crossing condition.

We consider the following game: The worker privately learns her type and chooses her education level e. Then the employer observe e but not θ, and make a wage offer to the worker through bargaining. Finally, the worker decides whether to accept. Assume that the outside option of both types of worker is zero and, to stick as closely to Spence (1973), the worker has all the bargaining power in the wage bargaining phase.[5] So the equilibrium wage is set equal to the employer's expected productivity of the worker conditional on observed education:

5) Note that if the employer has all the bargaining power in the bargaining, $w^*=0$ no matter what the worker's productivity is.

$$w(e) = \psi(\theta_H | e)\theta_H + (1 - \psi(\theta_H | e))\theta_L \in [\theta_L, \theta_H].$$

The employer still earns zero profit in expectation.

Again, we are interested in perfect prior-biased equilibria of this game:

1. The worker optimizes against the employer's strategy.
2. The employer draws inferences based on the posterior belief $\psi(\bullet | e)$ whenever possible.
3. The wage offer constitute a Nash equilibrium, given $\psi(\bullet | e) \forall e$

We focus on separating equilibria because, as mentioned above, there is no belief update and no change in pooling equilibria. A separating equilibrium is characterized by different types choosing different levels of education.

Claim 1

$$w^*(e^*(\theta_H)) = \psi(\theta_H|e)\theta_H + (1 - \psi(\theta_H|e))\theta_L$$
$$= ((1-a) \bullet 1 + a \bullet \lambda)\theta_H + (a - a \bullet \lambda)\theta_L < \theta_H$$

$$w^*(e^*(\theta_L)) = \psi(\theta_L|e)\theta_L + (1 - \psi(\theta_L|e))\theta_H$$
$$= ((1-a) \bullet 1 + a \bullet (1-\lambda))\theta_L + (a \bullet \lambda)\theta_H > \theta_L$$

Claim 1 directly follows from prior-biased inference (*). Even after observing $e^*(\theta_H)$, employer's prior-bias whispers to himself that there is a chance of being the low type. Consequently, the equilibrium wage gap between the types narrows due to the presence of employer's prior-bias.

Claim 2 $e^*(\theta_L) = 0.$

Otherwise the low type could increase her payoff by lowering education given the employer's posterior and wage offer. Therefore, the low type's welfare always improves under the prior-biased equilibria because of increased wage.

Claim 3 Separating equilibria can be Pareto-ranked.

The employer's expected profit is zero and the low type's utility is fixed throughout the equilibria by Claims 1 and 2, so the total welfare depends solely on the high type's utility. The equilibrium with the lowest education level for the high type is efficient and sometimes referred to as the Riley outcome.

For welfare comparison, we focus on the Riley outcome under Bayesian inference and that under prior-biased inference, and for tractability, let $c(e,\theta)=k_i e$, $i=L$, H, and $k_H<k_L$. Hence, we have

$$u(w,e,\theta)=w-c(e,\theta)=w-k_i e\,.$$

Claim 4 In the Riley outcome of perfect prior-biased equilibrium,

$$e^*(\theta_H)=\frac{(1-a)(\theta_H-\theta_L)}{k_L}\,.$$

Now we compare the welfares of Riley outcomes in both perfect Bayesian equilibrium and perfect prior-biased equilibrium. In the prefect Bayesian equilibrium, the high type's utility of the Riley outcome is

$$u_{Bayesian}(w^*,e^*,\theta_H)=\theta_H-k_H\frac{(\theta_H-\theta_L)}{k_L}$$

and the high type's utility of the Riley outcome in the perfect prior-biased equilibrium is

$$u_{prior-biased}(w^*, e^*, \theta_H) = (1-a+a\lambda)\theta_H + (a-a\lambda)\theta_L$$
$$-k_H\frac{(1-a)(\theta_H-\theta_L)}{k_L}.$$

Hence,

$$u_{prior-biased}(w^*, e^*, \theta_H) - u_{Bayesian}(w^*, e^*, \theta_H)$$
$$= a(\theta_H-\theta_L)\left(\frac{k_H}{k_L} - (1-\lambda)\right).$$

Claim 5 The utility of the high type worker in the perfect prior-biased equilibrium is greater than that in the perfect Bayesian equilibrium if $(1-\lambda)k_L \leq k_H < k_L$.

$(1-\lambda)k_L \leq k_H$ implies the proportion of low type workers $(1-\lambda)$ is not higher than the ratio of the education cost of the high type to the low type k_H/k_L. Then you may think k_H/k_L as the inverse of the high type's relative advantage in signaling over the low type.[6)]

Thus we have the following summary of welfare results: with the presence of prior-bias, (1) the equilibrium wage gap between two types is narrowed down, (2) wasteful education level decreases, (3) the low type worker always enjoys welfare improvement, and (4) the high type's welfare can be improved if the high type's (marginal) education cost is relatively high.

Ⅳ. Conclusion

Most economics and social science textbooks describe Bayes' rule as a basic principle of statistical inference and consider non-Bayesian updating to be an irrational bias which leads to sub-optimal outcome.

---

6) This interpretation was suggested by an anonymous referee.

In this paper, applying Lee, Lim and Zhao (2020)'s idea, we have illustrated the concept of prior-biased inference in the models of signaling. And we have drawn an interesting welfare implication that the receiver's prior-bias may work as a welfare-enhancing device even in the costly signaling environments. That is, contrary to the widespread beliefs that non-Bayesian inference is a sign of human irrationality and a social cost, we have shown that a behavioral bias can be beneficial.

Received: May 5, 2020. Revised: June 12, 2020. Accepted: June 19, 2020.

◈ *References* ◈

Benjamin, Daniel (2019), "Errors in Probabilistic Reasoning and Judgment Biases," In *Handbook of Behavioral Economics* (Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds.), Elsevier Press.

Bolton, Patrick and Mathias Dewatripont (2005), *Contract Theory*, MIT Press: Cambridge and London.

Charness, Gary and Chetan Dave (2017), "Confirmation Bias with Motivated Beliefs," *Games and Economic Behavior*, 104, 1-23.

Charness, Gary and Dan Levin (2005), "When Optimal Choices Feel Wrong: A Laboratory Study of Bayesian Updating, Complexity, and Affect," *American Economic Review*, 95, 1300-1309.

Cho, In-Koo, and David Kreps (1987). "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102, 179-221.

de Clippel, Geoffroy and Xu Zhang (2019), "Non-Bayesian Persuasion," Working paper, Brown University.

Eyster, Erik and Matthew Rabin (2005), "Cursed Equilibrium," *Econometrica*, 73(5), 1623-1672.

Epstein, Larry (2006), "An axiomatic Model of Non-Bayesian Updating," *Review of Economic Studies*, 73, 413-436.

Epstein, Larry, Jawwad Noor, and Alvaro Sandroni (2008), "Non-Bayesian Updating: A Theoretical Framework," *Theoretical Economics*, 3, 193-229.

Geller, Scott, and Gordon Pitz (1968), "Confidence and Decision Speed in the Revision of Opinion," *Organizational Behavior and Human Performance*, 3(2), 190-201.

Heifetz, Aviad, Chris Shannon, and Yossi Spiegel (2007), "What to Maximize if You Must" *Journal of Economic Theory*, 133, 31-57.

Lee, Yong-Ju, Wooyoung Lim, and Chen Zhao (2020), "Cheap Talk with Non-Bayesian Updating," Working paper.

Pitz, Gordon (1969), "An Inertia Effect (resistance to change) in the Revision of Opinion," *Canadian Journal of Psychology*, 23(1), 24.

Pitz, Gordon, Leslie Downing, and Helen Reinhold (1967), "Sequential Effects in the Revision of Subjective Probabilities," *Canadian Journal of Psychology*, 21(5), 381.

Spence, Michael (1973), "Job Market Signaling," *Quarterly Journal of Economics*, 87, 355-374.

Appendix: Equivalence of perfect prior-biased equilibrium and x-cursed equilibrium

Eyster and Rabin (2005), explaining a widely observed phenomenon of winner's curse, define a new equilibrium concept, called *x-cursed equilibrium*, which assumes that, with probability x, players in a Bayesian game underestimate the degree to which other players' actions are correlated with these other players' information. So, a cursed player incorrectly believes that each profile of types of the other players plays the same mixed action profile that corresponds to their average distribution of actions, rather than their true, type-specific action profile.[7] Although the motivation and interpretation of cursed equilibrium are completely different from the notion of perfect prior-biased equilibrium, we can easily show the mathematical equivalence of two equilibrium concepts in the signaling game environment. That is, we show that a prior-biased non-Bayesian in the signaling settings behaves just like a cursed Bayesian.

x-cursed equilibrium is defined in terms of a player's beliefs about others' actions as a function of types. Formally, a mixed-strategy profile σ is a *x-cursed equilibrium* if for each player i, $\theta_i \in \Theta_i$, and each y_i^* such that $\sigma_i(y_i^* | \theta_i) > 0$, $y_i^* \in \arg\max_{y_i \in Y_i} \sum_{\theta_{-i} \in \theta_{-i}} p_i(\theta_{-i} | \theta_i) \times$

7) Eyster and Rabin (2005, p1625) illustrate cursed equilibrium using a Akerlof's lemons model: "A buyer might purchase a used car from a seller at a price of \$1000. The seller knows whether the car is a lemon, worth \$0 to both, or a peach, worth \$3000 to the buyer and \$2000 to the seller. The buyer's prior is that each occurs equally. While a rational buyer would realize that the seller wishes to trade if and only if the car is a lemon and refuse to buy, a cursed buyer may buy the car. A x-cursed buyer believes that with probability x the seller sells with probability 1/2 irrespective of the type of car, so that the car being sold is a peach with probability $(1-x) \cdot 0 + x\frac{1}{2} = \frac{x}{2}$, and therefore worth $\frac{x}{2} \cdot 3000 = 1500x$. Hence a buyer cursed by $x > 2/3$ will buy the lemon car."

$\sum_{y_{-i}\in Y_{-i}}[x\overline{\sigma}_{-i}(y_{-i}|\theta_i)+(1-x)\sigma_{-i}(y_{-i}|\theta_{-i})]u_i(y;\theta)$, where Y_i denotes the set of player i's actions, Θ_i is the set of player i's types, player i's payoff function $u_i(y;\theta)$ depends on all players' actions and their types. $\overline{\sigma}_{-i}(y_{-i}|\theta_i)$ is the average strategy of other players, averaged over the other players' types and defined for each type of each player by $\overline{\sigma}_{-i}(y_{-i}|\theta_i)=\sum_{\theta_{-i}\in\theta_{-i}}p_i(\theta_{-i}|\theta_i)\ \bullet\ \sigma_{-i}(y_{-i}|\theta_{-i})$. When player i is of type θ_i, $\overline{\sigma}_{-i}(y_{-i}|\theta_i)$ is the probability that players $j\neq i$ play y_{-i} when they follow strategy σ_{-i}. Each player in the x-cursed equilibrium plays a best response to beliefs that with probability x her opponents play $\overline{\sigma}_{-i}(y_{-i}|\theta_i)$ regardless of their types, while with probability $1-x$ they play $\overline{\sigma}_{-i}(y_{-i}|\theta_i)$ depending on their types.

To show the equivalence, we invert the concept of x-cursed equilibrium in terms of a player's beliefs about others' types as a function of their actions played. Please note that the cursed players follow Bayes' rule. Let $\hat{\psi}_{\theta_i}(\theta_{-i}|y_{-i},\sigma_{-i})$ be type θ_i of player i's beliefs about other players' type θ_{-i} when they play action profile y_{-i} under strategy σ_{-i}. In a x-cursed equilibrium, from Bayes' rule,

$$
\begin{aligned}
&\hat{\psi}_{\theta_i}(\theta_{-i}|y_{-i},\sigma_{-i})\\
&=\frac{[x\overline{\sigma}_{-i}(y_{-i}|\theta_i)+(1-x)\sigma_{-i}(y_{-i}|\theta_{-i})]p_i(\theta_{-i}|\theta_i)}{\sum_{\theta'_{-i}\in\theta_{-i}}[x\overline{\sigma}_{-i}(y_{-i}|\theta_i)+(1-x)\sigma_{-i}(y_{-i}|\theta'_{-i})]p_i(\theta'_{-i}|\theta_i)}\\
&=\frac{[x\overline{\sigma}_{-i}(y_{-i}|\theta_i)+(1-x)\sigma_{-i}(y_{-i}|\theta_{-i})]p_i(\theta_{-i}|\theta_i)}{\overline{\sigma}_{-i}(y_{-i}|\theta_i)}\\
&=(1-x)\frac{\sigma_{-i}(y_{-i}|\theta_{-i})p_i(\theta_{-i}|\theta_i)}{\overline{\sigma}_{-i}(y_{-i}|\theta_i)}+xp_i(\theta_{-i}|\theta_i). \qquad (2)
\end{aligned}
$$

Then players choose the actions that maximize their expected payoff given these posterior beliefs, which constitute the x-cursed equilibrium. This posterior belief boils down to prior-biased posterior expressed in equation (1) in the signaling games in which there are two players and only the sender has private information. That is, two equilibrium concepts are equivalent if the sender's message is the single source of information as in the signaling games. In the language of prior-biased inference the first term of (2) is the Bayesian update of the common prior probability distribution and x indicates the degree of prior-bias.

Due to the mathematical equivalence, two equilibrium concepts share many common properties. Among them it is noteworthy that, as in our prior-biased equilibrium below, every pooling Bayesian Nash equilibrium is a x-cursed equilibrium for every x with a slightly different interpretation that in a pooling equilibrium, players' actions are independent of their types, so the relationship between others' actions and their information should be ignored.

신호, 사전적 정보에의 편향 그리고 후생

주 은 지* · 이 용 주**

논문초록

경제학을 포함해서 대부분의 사회과학 교과서에서 베이즈 규칙(Bayes' rule)을 사전적 정보와 추가적 정보를 결합하는 확률적 추론의 기본원리로 설명하고 있지만, 광범위한 실험연구들은 사람들의 정보처리 과정은 베이즈 규칙에서 체계적으로 벗어나 있음을 확인해주고 있다. 본 논문은 신호모형에서 수신자의 사전적 신념(prior beliefs)에 대한 편향이 경기자들의 전략적 인센티브와 후생적 함의에 어떤 영향을 미치는지 설명한다. 본 논문은 보편화된 예제들을 통해 비용이 드는 신호 환경에서 수신자의 사전적 신념에 대한 편향이 경기자들에게 후생증진 효과를 가져올 수 있음을 보인다. 즉, 베이지언 추론에서 벗어나는 것이 인간의 비합리성을 드러내는 것이고 사회적 비용이라는 일반적인 믿음과 달리, 이러한 행태적인 편향이 효용측면에서 이로운 것이 될 수 있음을 보인다.

주제분류 : B030200, B030904
핵심 주제 : 사전정보에 대한 편향, 신호, Beer-Quiche 게임, 구인시장에서의 신호

* 제1저자, 호서대학교 청소년문화 · 상담학과 교수, e-mail: jooej@hoseo.edu
** 교신저자, 영남대학교 경제금융학부 부교수, e-mail: yongjulee@ynu.ac.kr