WORKING PAPERS

N° 1244

September 2021

"A Note on Adverse Selection and Bounded Rationality"

Takeshi Murooka and Takuro Yamashita

Toulouse School of Economics

A Note on Adverse Selection and Bounded Rationality*

Takeshi Murooka† Takuro Yamashita‡

July 22, 2020

Abstract

We consider an adverse selection environment between an informed seller and an uninformed buyer, where no trade occurs when all buyers are rational. The buyer may be a "behavioral" type in the sense that he may take actions different from a rational type. We show that, for any incentive-feasible mechanism with any non-trivial trade, the buyer's ex-ante expected payoff is strictly negative. Our result implies that whenever trade occurs, some behavioral types *must* incur losses.

*We are grateful to Paul Heidhues as well as seminar participants at Munich and Toulouse for helpful comments. Takeshi acknowledges financial support from JSPS KAKENHI (JP16K21740, JP18H03640, JP19K01568, JP20K13451). Takuro acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 714693), and ANR under grant ANR-17-EURE-0010 (Investissements dAvenir program). All remaining errors are our own.

†Osaka School of International Public Policy, Osaka University. murooka@osipp.osaka-u.ac.jp

‡Toulouse School of Economics, University of Toulouse, Capitole. takuro.yamashita@tse-fr.eu

1 Introduction

Adverse selection is considered as one of the most important economic phenomena. Akerlof (1970) points out that, even if there exists positive gains of trade with probability one, essentially no trade is possible in equilibrium; Samuelson (1984) shows that such a no-trade outcome can be the only incentive-feasible allocation (both under certain parameter conditions).

However, there is accumulating evidence that some consumers are *behavioral* in the sense that they may make suboptimal decisions.[1] In the context of adverse selection, such behavioral agents may *improve* the total welfare. For example, Eyster and Rabin (2005) show that, in the environment where an informed seller trades with an uninformed buyer as in Akerlof (1970) and Samuelson (1984), if the uninformed buyer exhibits some kind of inferential naivety (namely, a buyer is "cursed"), then those two parties can potentially trade with positive probability. However, although the society in total can enjoy the positive trade surplus, the cursed buyer himself makes a loss (in terms of his "actual" payoff). Murooka and Yamashita (2020) extend this setting by allowing the buyer to be either the cursed type or the standard Bayesian-rational type, and show that not only the cursed type but also the rational type may be able to trade (even though the rational trader never trades in the absence of the cursed buyer). By definition, the rational type enjoys a non-negative expected payoff, while, as in Eyster and Rabin (2005), the cursed type makes a loss.

A natural question is whether this phenomenon is specific to a particular type of naivety (i.e., "cursedness") or is applicable to broader settings. To answer this question, this paper considers adverse selection with general types of such behavioral biases. The only assumption is that some consumer types may take actions that do not necessary optimize own "actual" payoffs, which encompasses virtually any type of biases including subjective probability, framing, model misspecification, random errors, and inferential naivety.

[1] See Heidhues and Kőszegi (2018) on consumer behavior in markets and Beshears, Choi, Laibson, and Madrian (2018) on household decision makings.

Our main result is the following. Consider an adverse-selection environment where only no-trade outcome is possible by rational agents. We show that, if there is any non-trivial trade under a mechanism which is incentive compatible for the informed seller, then the buyer's ex-ante expected payoff must be negative, no matter how the buyer behaves in the mechanism. As a corollary, because a rational type of the buyer always earns a non-negative expected payoff, some behavioral types of the buyer necessarily make negative expected payoffs in the mechanism.

The result may be interpreted as an impossibility result: In the adverse-selection environment, even though some trade may happen with behavioral types, it must be at the cost of these types' losses. Put it differently, the only mechanism that yields non-negative expected payoff to every type of the buyer is the no-trade mechanism.

On the other hand, in some situations, the society may accept some losses of the buyer, if it is compensated by a larger total surplus. Indeed, Murooka and Yamashita (2020) obtain incentive-feasible mechanisms where both rational and naive buyers trade even under severe adverse selection, attaining a strictly higher social surplus under certain conditions. In this sense, our theorem sheds light on a new trade-off between the social surplus and payoff losses of behavioral buyers.

2 Severe adverse selection with rational types

A seller has private information about the goods to be traded, denoted by $\theta \in \Theta$. The distribution of θ, denoted by $F \in \Delta(\Theta)$, is assumed to be common knowledge. A buyer has no private information. A (deterministic) trading outcome is denoted by $y \in Y$, which may include the information as to which goods are traded, the associated monetary transfers, and so on. The seller's ex-post payoff is denoted by $u_S(y, \theta)$, and the buyer's ex-post payoff is denoted by $u_B(y, \theta)$. The trading outcome includes a "no-trade outcome" $y = 0 \in Y$, and we assume $u_S(0, \theta) = u_B(0, \theta) = 0$ for normalization. A feasible allocation is a stochastic trading outcome, denoted by $x \in \Delta(Y)$. For each $i \in \{S, B\}$, let $u_i(x, \theta) = \int_y u_i(y, \theta) dx$ denote the expected payoff given x.

A fundamental observation in the literature (Akerlof, 1970; Samuelson, 1984) is that adverse selection can be so severe that only no-trade outcome is *incentive feasible* by rational traders (i.e., incentive compatible for the seller, and individually rational for both parties). Our goal is to obtain a different but related observation in case the buyer is not necessarily rational in the standard sense.

Specifically, consider an allocation rule $(x(\theta))_\theta$ that satisfy:

- (IC_S) incentive compatibility for the seller: for any θ, θ',

$$u_S(x(\theta), \theta) \geq u_S(x(\theta'), \theta),$$

- (IR_S) individual rationality for the seller: for any θ,

$$u_S(x(\theta), \theta) \geq 0,$$

- (IR_B) individual rationality for the buyer:

$$\int_\theta u_B(x(\theta), \theta) dF \geq 0.$$

Assumption 1 (Severe adverse selection)**.** An allocation rule $(x(\theta))_\theta$ satisfies (IC_S), (IR_S), and (IR_B) if and only if $x(\theta) = 0$ for (F-almost) all θ.

Example 1. The seller has an indivisible object. A deterministic trading outcome specifies whether the trade of the object occurs and the associated monetary transfer from the buyer to the seller. Both parties are risk neutral in monetary transfer, and hence, an allocation is identified by a pair $x = (q, p) \in [0, 1] \times \mathbb{R}$, where q represents the probability of trading the object and p represents the expected monetary transfer from the buyer to the seller.

The seller's ex post payoff is given by $u_S(q, p, \theta) = p - q\theta$, where $\theta \sim U(0, 1)$ can be interpreted as the seller's opportunity cost of trading. The buyer's ex post payoff is $u_B(q, p, \theta) = q\alpha\theta - p$, where $\alpha\theta$ can be interpreted as the buyer's valuation for the object. Assume $\alpha \in (1, 2)$.

By the standard argument based on the envelope theorem, the combination of (IC_S) and (IR_S) implies:

$$p(\theta) \geq q(\theta)\theta + \int_\theta^1 q(z)dz.$$

Then, the buyer's expected payoff is at most:

$$\begin{aligned} & \int_0^1 \left(q(\theta)\alpha\theta - q(\theta)\theta - \int_\theta^1 q(z)dz \right) d\theta \\ = & \int_0^1 q(\theta)(\alpha - 2)\theta d\theta, \end{aligned}$$

which is negative unless $q(\theta) = 0$ for almost all θ.

3 Severe adverse selection with behavioral types

Eyster and Rabin (2005) and Murooka and Yamashita (2020) consider the situation where the buyer may be "cursed" in that he under-appreciates the correlation between the seller's private information and action. They show that (i) non-trivial trade (and hence a positive social surplus) can occur; while (ii) the cursed buyer makes a loss.

Although those two papers consider very specific behavioral types, we show that the same property holds under any kind of behavioral types. Namely, some buyer type must make a loss unless it is a no-trade mechanism.

To formally state our main result, let $k = 1, \ldots, K$ denote the buyer's behavioral type, where $g_k \in (0, 1)$ denotes the probability of each type k. Each type k's ex-post payoff is given by $u_B(x, \theta)$, invariant with respect to k. This means that the buyer's behavioral type is purely about his action pattern. We assume that θ and k are independent, which means that the buyer's behavioral type is not informative about the seller's value.[2] Let $x(\theta, k)$ denote the allocation if the agents' types are (θ, k). For the seller, we require:

- (IC_S) incentive compatibility for the seller: for any θ, θ',

$$\sum_k g_k u_S(x(\theta, k), \theta) \geq \sum_k g_k u_S(x(\theta', k), \theta),$$

- (IR_S) individual rationality for the seller: for any θ,

$$\sum_k g_k u_S(x(\theta, k), \theta) \geq 0.$$

We say that buyer type k makes a loss if $\int_\theta u_B(x(\theta, k), \theta) dF < 0$.

Theorem 1. Suppose that an allocation rule $(x(\theta, k))_{(\theta,k)}$ satisfies (IC_S) and (IR_S). Then, unless $x(\theta, k) = 0$ for all k and (F-almost) all θ, some buyer type makes a loss.

Proof. We show a slightly stronger claim: If an allocation rule $(q(v, k), p(v, k))_{(v,k)}$ satisfies (IC_S), (IR_S), and makes the buyer's *ex ante* expected payoff (i.e., the buyer's expected payoff before realizing own type) non-negative, then $q(v, k) = 0$ for all k and (F-almost) all v.

Specifically, the buyer's ex ante expected payoff is:

$$\sum_k g_k \left(\int_\theta u_B(x(\theta, k), \theta) dF \right).$$

Define $x(\theta) = \sum_k g_k x(\theta, k)$ for each θ. Then, (IC_S) becomes:

$$u_S(x(\theta), \theta) \geq u_S(x(\theta'), \theta),$$

[2] We think this is a natural assumption. However, generalization to the correlated case is also possible, as long as we strengthen the seller's incentive compatibility and individual rationality to their ex-post version. This strengthening is in order to avoid a Cremer-McLean type mechanism (Crémer and McLean, 1988), which extracts the seller's information rent by asking the seller to bet on the buyer's type realization.

(IR_S) becomes:

$$u_S(x(\theta), \theta) \geq 0,$$

and the buyer's ex ante expected payoff becomes:

$$\int_v u_B(x(\theta), \theta) dF.$$

Therefore, by Assumption 1, if the buyer's unconditional expected payoff is non-negative, we must have $x(\theta) = 0$ for (F-almost) all θ, implying $x(\theta, k) = 0$ for all k and for (F-almost) all θ. $\square$

Theorem 1 points out a fundamental problem of an adverse selection environment. Namely, whatever decision rules of the buyer one introduces in order to avoid the no-trade outcome, such avoidance is possible only at the risk of some buyer types, or more precisely, at the risk of the buyer's ex ante payoff.

Notice that we have little restriction on which kinds of behavioral types we consider. It includes any type of misinference (Eyster, 2019), inattention (Gabaix, 2019), random errors such as Goeree, Holt, and Palfrey (2008) and Gabaix, Laibson, Li, Li, Resnick, and de Vries (2016), or heterogeneous priors for the value. There may exist multiple kinds of behavioral types coexisting, with possibly different degrees. Theorem 1 is also agnostic about the class of mechanisms and the equilibrium concept, as long as the seller is best-responding. Such non-dependence on mechanisms allows for analogy-based expectation, framing-based biases, and model misspecifications such as Jehiel (2005), Ahn and Ergin (2010), Esponda and Pouzo (2016), and Spiegler (2016).

References

AHN, D. S., AND H. ERGIN (2010): "Framing Contingencies," *Econometrica*, 78(2), 655–695.

Akerlof, G. A. (1970): "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84(3), 488–500.

Beshears, J., J. J. Choi, D. Laibson, and B. C. Madrian (2018): "Behavioral Household Finance," in *Handbook of Behavioral Economics: Foundations and Applications 1*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, vol. 1, chap. 3, pp. 177–276. North Holland.

Crémer, J., and R. P. McLean (1988): "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56(6), 1247–1257.

Esponda, I., and D. Pouzo (2016): "Berk-Nash Equilibrium: A Framework for Modeling Agents with Misspecified Models," *Econometrica*, 84(3), 1093–1130.

Eyster, E. (2019): "Errors in Strategic Reasoning," in *Handbook of Behavioral Economics: Foundations and Applications 2*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, vol. 2, chap. 3, pp. 187–259. North Holland.

Eyster, E., and M. Rabin (2005): "Cursed Equilibrium," *Econometrica*, 73(5), 1623–1672.

Gabaix, X. (2019): "Behavioral Inattention," in *Handbook of Behavioral Economics: Foundations and Applications 2*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, vol. 2, chap. 4, pp. 261–343. North Holland.

Gabaix, X., D. Laibson, D. Li, H. Li, S. Resnick, and C. G. de Vries (2016): "The Impact of Competition on Prices with Numerous Firms," *Journal of Economic Theory*, 165, 1–24.

Goeree, J. K., C. A. Holt, and T. R. Palfrey (2008): "Quantal Response Equilibrium," in *Palgrave Dictionary of Economics*, pp. 1–11.

Heidhues, P., and B. Kőszegi (2018): "Behavioral Industrial Organization," in *Handbook of Behavioral Economics: Foundations and Applications 1*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, vol. 1, chap. 6, pp. 517–612. North Holland.

Jehiel, P. (2005): "Analogy-based expectation equilibrium," *Journal of Economic Theory*, 123(2), 81–104.

Murooka, T., and T. Yamashita (2020): "Optimal Trade Mechanisms with Adverse Selection and Inferential Mistakes," Working Paper.

Samuelson, W. (1984): "Bargaining under Asymmetric Information," *Econometrica*, 52(4), 995–1005.

Spiegler, R. (2016): "Bayesian Networks and Boundedly Rational Expectations," *Quarterly Journal of Economics*, 131(3), 1243–1290.