

DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

STATUS QUO BIAS IN BARGAINING: AN EXTENSION OF THE
MYERSON SATTERTHWAITTE THEOREM WITH AN APPLICATION TO
THE COASE THEOREM

Richard D. McKelvey

Talbot Page



SOCIAL SCIENCE WORKING PAPER 1035

April 2000

Status Quo Bias in Bargaining: An extension of the Myerson Satterthwaite Theorem with an application to the Coase Theorem*

Richard D. McKelvey
California Institute of Technology

Talbot Page
Brown University

April 7, 2000

*This paper was previously titled "The Coase Theorem with Private Information." The financial support of the National Science Foundation (Grant #SBR-9223701) is gratefully acknowledged. We thank Kim Border, John Duggan, Tom Palfrey, Andy Postlewaite, and Harl Ryder for valuable discussions and suggestions, and especially Kim Border for pointing us to the right duality results necessary in the proof of part B of our proposition.

Status Quo Bias in Bargaining: An extension of the Myerson Satterthwaite Theorem with an application to the Coase Theorem

Richard D. McKelvey

Talbot Page

Abstract

We use a generalized version of the Myerson-Satterthwaite theorem to study inefficiencies in bilateral bargaining over trade of an indivisible good, where there is two sided private information on the valuations. We show that when preferences are convex and quasi linear, and when the private information represents the magnitude of the utility gain or loss and follows a uniform distribution, that the most efficient mechanism always exhibits a bias towards the status quo. In the case that utility functions are quadratic in the amount traded, we prove that for any incentive compatible direct mechanism, there is an expected bias towards the disagreement point. In other words, for the class of preferences we study, there is a strategic advantage to property rights in the Coase bargaining setup in the presence of incomplete information.

1 Introduction

We consider the problem of bilateral trade of a divisible good, when there is two sided incomplete information about the valuations. The questions we ask are whether in such a case there will be inefficiencies, and if so, whether there is any bias in the direction that outcomes tend to deviate from the efficient allocation. This is of course not a new problem. The Myerson-Satterthwaite (MS) theorem and its corollary identify conditions under which bilateral trading leads to inefficiency. In this paper we generalize the MS theorem, investigate the existence of bias when there is no efficient equilibrium, and apply the results to the Coase theorem.

Generalization of the MS Theorem. The original MS theorem is for an indivisible good, or equivalently for the case where the good is divisible but utilities are linear. Theorems 1 and 3 generalize the MS theorem to non-linear utilities and a divisible good. The generalization is straightforward and follows closely the proof of the original theorem.

Bias. Theorem 4 shows for concave utilities and separable information, that if there is no fully efficient Bayesian Equilibrium (BE), the most efficient BE has a bias towards the status quo: the equilibrium amount traded will never be more than the efficient amount of trade and will sometimes be less. In Proposition 5, we show that if utilities are quadratic, that for every BE, the expected amount of trade is less than the expected efficient amount of trade.

Application to the Coase Theorem. The Coase theorem says that when property rights are well defined and enforced and when there are no transaction costs, direct bargaining over the amount of an environmental harm will lead to (A) efficiency and (B) neutrality (the amount of the harm will be the same no matter which way the rights were initially allocated). As noted by Fudenberg and Tirole (1992, p 279), the original MS theorem is in apparent conflict with (A), the efficiency claim of the Coase theorem. The generalization of the MS theorem suggests that the conflict continues to

exist when there are non-linear utilities and a divisible harm,¹ and our investigation of the bias suggests a conflict with (B), the neutrality claim. For example, when rights are allocated to the victim, strategic use of private information appears to lead toward less pollution than is efficient, and when rights are allocated toward the polluter strategic use of private information appears to lead to less abatement (and hence more pollution) than is efficient.

These results imply that the Coase theorem does not extend to the environment of private information. In a separate experimental paper (McKelvey and Page 1997) we estimate the amount of inefficiency and the strength of the bias toward the disagreement point in a specific game form. We find that both are substantial and the bias is in the predicted direction.

Other literature has considered the effects of private information in the Coase Theorem. Samuelson [1985] argues that in the indivisible good case that inefficiencies can be expected. Farrell [1987] considers a case of a divisible good with quadratic utility and argues that in the presence of private information that one can achieve efficient incentive compatible mechanisms, but not if voluntary participation is required. Illing [1992] considers a divisible good with quadratic utility similar to the one considered in section 4 of this paper, and shows that under a particular game form, that inefficiency arises and distorts the outcome in the direction of the status quo. The above papers are for bilateral bargaining. Mailath and Postlewaite [1990] show that for n -person bargaining over an indivisible good, that the probability of an efficient action goes to zero. We only consider bilateral bargaining. Our contribution to the literature on the Coase Theorem is to show that for the case of an indivisible good, at least for the quadratic case considered in the literature, inefficiencies can be expected to occur for any Bayes equilibrium, regardless of the game form that is used, and that these inefficiencies have

¹Contrast this with results of Crampton, Gibbons, and Klemperer (1987), who find that there exist ex post efficient mechanisms for dissolving a partnership, as long as the ownership shares of each individual are not too great. In the Coase setting one of the agents has an ownership share of one. So there is no conflict between these results.

a bias towards the status quo.

2 Generalization of Myerson-Satterthwaite

There are two agents, a *buyer* and *seller*, and a set of outcomes $Z = X \times Y$, where $X = [0, x_H]$ and $Y = [0, y_H]$ are each closed intervals in \mathfrak{R} . Points in Z are denoted (x, y) , with $x \in X$ and $y \in Y$; x is the amount of the good sold, and y is the payment, or transfer from buyer to seller.² Each agent's utility is a function of the outcome, and of their own privately observed type, $b \in B = [b_L, b_H]$ for the buyer, and $c \in C = [c_L, c_H]$ for the seller. Utility functions, $u : Z \times B \mapsto \mathfrak{R}$ and $v : Z \times C \mapsto \mathfrak{R}$, for the buyer and seller, respectively, take the form

$$u(x, y; b) = f(x, b) - y \quad \text{Buyer's payoff} \quad (1)$$

$$v(x, y; c) = y - g(x, c) \quad \text{Seller's payoff} \quad (2)$$

$$b \sim P[b_L, b_H]$$

$$c \sim Q[c_L, c_H] \quad \text{Independent random variables} \quad (3)$$

$$(x, y) = (0, 0) \quad \text{Status quo, or disagreement point} \quad (4)$$

where f and g are twice continuously differentiable real valued functions on $X \times B$, and $X \times C$, respectively, satisfying $f(0, b) = g(0, c) = 0$ for all b, c , and P and Q are the cumulative density functions of absolutely continuous measures on the space of types of each agent. We let p and q be the corresponding density functions of P and Q .

The two agents attempt to come to an agreement on a point $(x, y) \in Z$. If they fail to agree, they get the disagreement point $(0, 0)$. We do not specify a particular extensive form game. Rather we characterize the allocations that can arise as a Bayesian equilibrium to any extensive form game. Using the revelation principle, it is sufficient to

²Alternatively, in the Coasian context studied later, under victim's rights, the buyer can be thought of as the polluter, and the seller as the victim, x the amount of pollution, and y the transfer. Under polluter's rights, the buyer is the victim, the seller is the polluter, and x the amount of abatement.

consider direct mechanisms, of the form $(x, y) = (x(b, c), y(b, c)) : B \times C \mapsto Z$, and to characterize the set of incentive compatible individually rational direct mechanisms.

The Myerson-Satterthwaite theorem [1983, theorem 1] characterizes the BE for bilateral trade of an indivisible good with incomplete information about the valuations. We are interested in an environment where goods are divisible and preferences of individuals need not be linear over the amount of the good. So we begin by proving an extended version of the original MS result that applies to this environment.

For any b, c and x define:

$$\text{social surplus} \quad s(x; b, c) = f(x, b) - g(x, c) \quad (5)$$

$$\text{efficient allocation} \quad x^*(b, c) = \arg \max_{x \in X} s(x; b, c) \quad (6)$$

We restrict attention to the set \mathcal{X} of integrable functions on $T = B \times C$. For any $x, y \in \mathcal{X}$, we define

$$\tilde{U}(b) = E_c [f(x(b, c), b) - y(b, c)] \quad (7)$$

and

$$\tilde{V}(c) = E_b [y(b, c) - g(x(b, c), c)]. \quad (8)$$

Define $M : \mathcal{X} \mapsto \Re$ by

$$M(x) = E \left[\left(f(x(b, c), b) - \frac{(1 - P(b))}{p(b)} f_2(x(b, c), b) \right) - \left(g(x(b, c), c) + \frac{Q(c)}{q(c)} g_2(x(b, c), c) \right) \right] \quad (9)$$

Our generalized version of the MS theorem is split into two parts, a preliminary proposition, and a characterization theorem. The generalization to non-linear utility and a divisible good goes through straightforwardly and follows very closely the original proof of Myerson and Satterthwaite.

Theorem 1 (*Generalization of MS theorem, part 1*) *Assume individual utility functions and information are as in (1) - (4). Then any incentive compatible direct mechanism, $(x(b, c), y(b, c))$ satisfies*

$$M(x) = \tilde{U}(b_L) + \tilde{V}(c_H) \quad (10)$$

The proof of this result, as well as the proof of 3 below follow closely along the lines of the original Myerson-Satterthwaite [1983] theorem, with the main difference being the added complication of the f and g functions. Because of the differences, the proofs are presented in the appendix.

Corollary 2 *If $M(x^*) < 0$, there is no incentive compatible individually rational mechanism which is ex post efficient.*

Proof. Since the status quo is $(x, y) = (0, 0)$, by IR, $\tilde{U}(b_L) \geq u(0, 0, b_L) = f(0, b_L) = 0$ and $\tilde{V}(c_H) \geq v(0, 0, c_H) = g(0, c_H) = 0$. Thus by Theorem 1, for a mechanism to satisfy IC and IR, we must have $M(x) = \tilde{U}(b_L) + \tilde{V}(c_H) \geq 0$. Thus if $M(x^*) < 0$, there is no ex post efficient IR, IC mechanism. ■

Theorem 3 *(Generalization of MS, Theorem 1, part 2) Assume individual utility functions and information are as in (1) - (4). Then for any $x \in \mathcal{X}$ a sufficient condition for there to exist a $y \in \mathcal{X}$ such that (x, y) is a direct mechanism satisfying IC and interim IR is*

$$M(x) \geq 0 \tag{11}$$

$$F_{21} = F_{12} \geq 0, \frac{\partial}{\partial t} F_2(t, t) \geq 0, \tag{12}$$

$$G_{21} = G_{12} \leq 0, \frac{\partial}{\partial t} G_2(t, t) \leq 0. \tag{13}$$

where $F(\hat{b}, b) = E_c[f(x(\hat{b}, c), b)]$ and $G(\hat{c}, c) = E_b[g(x(b, \hat{c}), c)]$. If $f(x, b) = b\bar{f}(x)$, and $g(x, c) = c\bar{g}(x)$, for some \bar{f} and \bar{g} that do not depend on b or c , then the above conditions are also necessary

Note that Corollary 2 is the analogue of the MS Corollary 1.

Example 1: $f(x, b) = bx$, $g(x, c) = cx$ and $x_H = 1$.

For this case the above Theorems reduce to the original MS Theorem, parts (1) and (2).

Example 2: $X = Y = [0, 1]$, $B = [0, b_H]$, $C = [0, c_H]$ and

$$\begin{aligned} f(x, b) &= b(2x - x^2) \\ g(x, c) &= cx^2, \end{aligned} \tag{14}$$

with uniform priors over B and C (i. e., $p(b) = 1/b_H$ and $q(c) = 1/c_H$).

This example is analysed in greater detail in section 4, where we look at implications of the above results for the Coase theorem under quadratic utilities.

The social surplus is $s(x; b, c) = 2bx - (b + c)x^2$. For any b, c , this is maximized at $x^*(b, c) = \frac{b}{b+c}$, which is the efficient allocation.

For this second example, using that $f(x, b) = b f_2(x, b)$ and $g(x, c) = c g_2(x, c)$, and substituting in for the remaining terms,

$$\begin{aligned} M(x) &= E[(2b - b_H) f_2(x(b, c), b) - 2c g_2(x(b, c), c)] \\ &= E[(2b - b_H)(2x - x^2) - 2cx^2], \end{aligned}$$

and setting $x^* = \frac{b}{b+c}$,

$$M(x^*) = E \left[\frac{b_H b^2}{(b+c)^2} - 2 \frac{b(b_H - b)}{b+c} \right].$$

So

$$\begin{aligned} b_H c_H M(x^*) &= \int_0^{b_H} \int_0^{c_H} \left[\frac{b_H b^2}{(b+c)^2} - 2 \frac{b(b_H - b)}{b+c} \right] db dc \\ &= \int_0^{b_H} \left(-\frac{b_H b^2}{b+c} - 2b(b_H - b) \ln(b+c) \right) \Big|_0^{c_H} db \\ &= \int_0^{b_H} \left[\frac{b b_H c_H}{b+c_H} + 2b(b_H - b) \ln \left(\frac{b}{b+c_H} \right) \right] db \\ &= \frac{1}{3} \left(b^2 c_H - 2b c_H^2 + b^2 (3b_H - 2b) \ln \left(\frac{b}{b+c_H} \right) + 2c_H^3 \ln(b+c_H) \right) \Big|_0^{b_H} \\ &= \frac{1}{3} \left(b_H^2 c_H - 2b_H c_H^2 + b_H^3 \ln \left(\frac{b_H}{b_H+c_H} \right) - 2c_H^3 \ln \left(\frac{c_H}{b_H+c_H} \right) \right). \end{aligned} \tag{15}$$

Define $W(b, c) = bcM(x^*)$. Then for any $0 < b_H$, and $0 < c_H$, $M(x^*) < 0 \Leftrightarrow W(b_H, c_H) < 0$. But $W(b, c) = 0$ if either b or c is zero. Setting $t = \frac{c}{b}$,

$$\begin{aligned} W_1(b, c) &= b^2 \ln \left(\frac{b}{b+c} \right) + b^2 \frac{c}{b+c} < 0 \\ &\Leftrightarrow -\ln(1+t) + \frac{t}{1+t} < 0. \end{aligned} \tag{16}$$

Similarly, setting $t = \frac{b}{c}$,

$$\begin{aligned} W_2(b, c) &= -bc \frac{b+2c}{b+c} - 2c^2 \ln \left(\frac{c}{b+c} \right) < 0 \\ &\Leftrightarrow -\left(\frac{t}{t+1} \right) \left(\frac{t+2}{2} \right) + \ln(1+t) < 0 \end{aligned} \tag{17}$$

But using the fact that for any $t > 0$,

$$\frac{t}{1+t} < \ln(1+t) < \left(\frac{t}{t+1} \right) \left(\frac{t+2}{2} \right),$$

it follows that $W_1(b, c) < 0$ and $W_2(b, c) < 0$ for all $b, c > 0$. Hence, it follows that $W(b_H, c_H) < 0$ for all $b_H, c_H > 0$. Therefore, $M(x^*) < 0$ for all $b_H, c_H > 0$.³ So from Corollary 1, it follows that for the quadratic model, there is no incentive compatible direct mechanism that is efficient.

3 Bias of the maximally efficient mechanism

This section considers maximally efficient mechanisms, and explores the direction outcomes tend to deviate from the efficient allocation, or the “bias” in these mechanisms. Gresik (1991) characterizes maximally efficient mechanisms for the case of indivisible goods, and shows that ex ante maximally efficient mechanisms are not ex post efficient. His result is for the Myerson Satterthwaite setup of indivisible goods, and hence does not address the question of bias.

³In an earlier version of this paper, we incorrectly computed that $M(x^*) > 0$ for some b_H, c_H . and hence incorrectly claimed that this provided a counter-example to the possibility of extending the Myerson Satterthwaite Corollary to a quadratic environment.

We consider only the case where $f(x, b) = b\bar{f}(x)$, and $g(x, c) = c\bar{g}(x)$ for some \bar{f}, \bar{g} . We assume that the distribution of b and c are both uniform over $[0, 1]$. In other words, $P(b) = b$, and $Q(c) = c$. Since we assumed that $f(x, b) = g(x, c) = 0$ for all b, c , it follows that $\bar{f}(0) = 0$ and $\bar{g}(0) = 0$. We make the additional assumptions that $\bar{f}'(x) \geq 0$, with strict inequality for $x < x_H$, $\bar{g}'(x) \geq 0$, with strict inequality for $x > 0$, and $\bar{f}'' < 0$, $\bar{g}'' > 0$.

Under the above assumptions, $F_2(\hat{b}, b) = \bar{F}(\hat{b})$, and $G_2(\hat{c}, c) = \bar{G}(\hat{c})$, where $\bar{F}(\hat{b}) = E_c[\bar{f}(x(\hat{b}, c))]$, and $\bar{G}(\hat{c}) = E_b[\bar{g}(x(b, \hat{c}))]$. Hence, from Theorem 3, it follows that for any $x \in \mathcal{X}$, necessary and sufficient conditions that there exists a $y \in \mathcal{X}$ such that (x, y) is a direct mechanism satisfying incentive compatibility and individual rationality is that

$$M(x) \geq 0, \tag{18}$$

$$\bar{F} \text{ is non decreasing,} \tag{19}$$

$$\bar{G} \text{ is non increasing.} \tag{20}$$

Therefore, we define a *maximally efficient mechanism* to be any $x^o \in \mathcal{X}$ satisfying

$$\begin{aligned} x^o &\in \arg \max_{x \in \mathcal{X}} E[s(x; b, c)] \\ \text{s.t.} \quad &M(x) \geq 0, \bar{F}' \geq 0, \bar{G}' \leq 0 \end{aligned}$$

Theorem 4 (Bias) *If $M(x^*) < 0$ (No fully efficient BE), and x^o is a maximally efficient mechanism, and $0 < x^*(b, c) < x_H$ on a set of positive measure, then $x^o(b, c) \leq x^*(b, c)$ a. e., with $x^o(b, c) < x^*(b, c)$ on a set of positive measure.*

4 Bias for the quadratic case

We now consider in more detail the quadratic case introduced in equation (14) of the second section. Recall that for the quadratic case, we assumed that $X = Y = [0, 1]$, $B = [0, b_H]$, $C = [0, c_H]$ and

$$f(x, b) = b(2x - x^2)$$

$$g(x, c) = cx^2, \quad (21)$$

with uniform priors over B and C . Hence, we can write

$$\begin{aligned} u(x, y; b) &= b(2x - x^2) - y && \text{Polluter's payoff} \\ v(x, y; c) &= y - cx^2 && \text{Victim's payoff} \end{aligned} \quad (22)$$

The quadratic case has been used in the literature on the Coase Theorem ⁴. In this interpretation, we can think of the buyer as the *polluter*, the seller as the *victim*, and the property rights as being assigned to the victim. Property rights in this framework are represented by the location of the disagreement point, or status quo. We have assumed that the disagreement point is at the point $(0, 0)$, which is at the value of $x = 0, y = 0$ corresponding to the ideal point that can be enforced by the victim, if the victim can choose the level of x .

To consider the case where property rights are assigned to the polluter, we would change the disagreement point to the point $(1, 0)$ where $x = x_H = 1$, corresponding to the ideal point of the polluter. But with a lump sum tax of b to the polluter, and a lump sum subsidy of c to the victim (which does not change the Bayes equilibria of the game), we get that

$$\begin{aligned} u(x, y; b) &= b(2x - x^2) - y - b &= t - bz^2 &= v(z, t; b) \\ v(x, y; c) &= y + c - cx^2 &= c(2z - z^2) - t &= u(z, t; c) \end{aligned} \quad (23)$$

where $z = 1 - x$ is the amount of abatement, and $t = -y$ is the transfer from victim to polluter. Setting the disagreement point to be the point $z = 0, t = 0$. we see that the case where the property rights are assigned to the polluter is formally equivalent to the case of victim rights, were we simply reverse the roles of the victim and the polluter.

Because of this symmetry of the two rights assignments, it is only necessary to consider one case. If we show that there is a bias towards the disagreement point in

⁴Similar models have been used by Coase [1960], Turvey [1963], Demsetz [1966], and Mishan [1971]. Coase, in his table and discussion of the rancher and farmer, uses a quadratic cost function for the victim (farmer) and side payments (pp. 111-116); Turvey (p. 311, Figure 1), Demsetz (p. 68, Figure 1), and Mishan (p. 20, Figure 2) have quadratic cost and benefit functions of pollution, and side payments. However, none of these models explicitly specify the information structure.

the case of victim rights, polluter rights will lead to a corresponding bias towards the disagreement point that can be enforced with polluter rights. Hence, a status quo bias in the framework of the quadratic model of equation (22) implies that there is a bias of the equilibrium level of pollution in the direction of the holder of the property rights.

In the quadratic case, we are able to show that *all* BE (not just the maximally efficient one) have a bias towards the status quo, yielding a strategic advantage to property rights:

Proposition 5 *Assume $X = Y = [0, 1]$, with individual utility functions as in (22), and that $B = C = [0, 1]$, with uniform priors on B and C . Then any incentive compatible direct mechanism $(x, y) \in \mathcal{X} \times \mathcal{X}$ satisfying individual rationality satisfies:*

(A) *It is not ex post efficient: $x(b, c) \neq x^*(b, c) = \frac{b}{b+c}$ on some set of positive measure.*

(B) *$E[x(b, c)] < E[\frac{b}{b+c}]$.*

We can interpret these as implications about the allocations that can arise in the Bayes equilibria (BE) of any bargaining game with quadratic payoffs as in (14). Then (A) says there is inefficiency for any BE. (B) says that on average, x will be less than the efficient amount. Thus (B) gives a direction to the deviations of outcomes from efficient allocations proven in (A): the direction of the deviations from efficiency is toward the disagreement point $x = 0$.

5 References

- Aliprantis, Charalambos D., and Kim C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, (Berlin: Springer Verlag, 1994).
- Coase, Ronald, "The problem of social cost," *Journal of Law and Economics*, 3 (1960): 1-31.
- Crampton, P., R. Gibbons, and P. Klemperer, "Dissolving a partnership efficiently," *Econometrica*, 615, (1987): 615-632.
- Farrell, Joseph, "Information and the Coase Theorem," *Economic Perspectives*, 1 (1987): 113-129.
- Demsetz, Harold, "Some aspects of property rights," *Journal of Law and Economics*, October, (1966): 61-70.
- Fudenberg, D., and J. Tirole, *Game Theory*, (Cambridge, MIT Press, 1991).
- Gresik, T. A., "*Ex ante* efficient, *ex post* individually rational trades," *Journal of Economic Theory*, 131, (1991): 131-145.
- Hurwicz, Leonid. "What is the Coase theorem," *Japan and the World Economy*, 7 (1995): 49-74.
- Illing, G., "Private information as transaction costs: The Coase Theorem revisited," *Journal of Institutional and Theoretical Economics*, 148 (1992): 558-576.
- Krishna, V., and M. Perry, "Efficient mechanism design," *mimeo*, (1997).
- Luenberger, D. G., *Optimization by Vector Space Methods*, (New York: John Wiley and Sons, 1969).

- Mailath, G., and A. Postlewaite, "Asymmetric-information bargaining problems with many agents," *Review of Economic Studies*, 57 (1990): 351-368.
- McKelvey, R., and T. Page, "An experimental study of the Coase Theorem with incomplete information", *Caltech Working Paper*, April, 1997.
- Mishan, E. J., "The postwar literature on externalities: An interpretive essay," *Journal of Economic Literature*, (1971): 1-28.
- Myerson, R. B., and M. A. Satterthwaite, "Efficient mechanisms for bilateral trading," *Journal of Economic Theory*, 29 (1983): 265-281.
- Rob, R., "Pollution claims settlements with private information," *Journal of Economic Theory*, (1989): 307-333.
- Samuelson, William, "A comment on the Coase Theorem" in *Game-Theoretic Models of Bargaining*, edited by A. Roth, (New York: Cambridge University, 1985).
- Turvey, Ralph, "On divergences between social cost and private cost," *Economica*, (1963): 309-13.

APPENDIX

This appendix presents proofs of the main results of the paper.

5.1 Proof of Theorem 1

Proof. The idea is to derive two expressions for the expected social surplus and equate them. From (5), the expected social surplus is

$$E[s(x(b, c); b, c)] = E[f(x(b, c), b) - g(x(b, c), c)] \quad (24)$$

To obtain the second measure of social surplus, first define $U(\hat{b}, b)$ to be agent 1's expected utility from reporting \hat{b} when his true type is b . Thus,

$$U(\hat{b}, b) = F(\hat{b}, b) - E_c[y(\hat{b}, c)] \quad (25)$$

where $F(\hat{b}, b) = E_c[f(x(\hat{b}, c), b)]$. Note from (7), that $\tilde{U}(b) = U(b, b)$. Thus

$$\tilde{U}'(b) = U_1(b, b) + U_2(b, b) = F_2(b, b),$$

since IC requires, for $b \in (b_L, b_H)$, $U_1(b, b) = 0$. Thus we have,

$$\tilde{U}(b) = \int_{b_L}^b F_2(t, t) dt + \tilde{U}(b_L) \quad (26)$$

Taking the expectation, and reversing the order of integration yields

$$\begin{aligned} E_b[\tilde{U}(b)] &= \int_B \int_{b_L}^b F_2(t, t) dt p(b) db + \tilde{U}(b_L) = \int_B F_2(t, t) \left[\int_t^{b_H} p(b) db \right] dt + \tilde{U}(b_L) \\ &= \int_B (1 - P(b)) F_2(b, b) db + \tilde{U}(b_L). \end{aligned}$$

A similar argument for agent 2 shows,

$$\tilde{V}(c) = \int_c^{c_H} G_2(t, t) dt + \tilde{V}(c_H), \quad (27)$$

and

$$E_c[\tilde{V}(c)] = \int_C \int_c^{c_H} G_2(t, t) dt dc + \tilde{V}(c_H) = \int_C Q(c) G_2(c, c) dc + \tilde{V}(c_H).$$

Thus, a second measure of the expected social surplus is the sum of these two expected earnings

$$\begin{aligned}
& E_b[\tilde{U}(b)] + E_c[\tilde{V}(c)] \\
&= \int_B (1 - P(b)) F_2(b, b) db + \int_C Q(c) G_2(c, c) dc + \tilde{U}(b_L) + \tilde{V}(c_H) \\
&= \int_C \int_B [(1 - P(b)) q(c) f_2(x(b, c), b) \\
&\quad + Q(c) p(b) g_2(x(b, c), c)] db dc + \tilde{U}(b_L) + \tilde{V}(c_H). \tag{28}
\end{aligned}$$

Equating 24 with 28, collecting terms, and factoring out the $p(b) q(c)$, gives $M(x) = \tilde{U}(b_L) + \tilde{V}(c_H)$. ■

5.2 Proof of Theorem 3

Proof. (*Sufficiency:*) Pick y so that

$$y(b, c) = \int_{b_L}^b F_1(t, t) dt + \int_{c_L}^c G_1(t, t) dt + F(b_L, b_L) - \int_C [1 - Q(t)] G_1(t, t) dt$$

Since only the first term depends on b , it follows that

$$y(b, c) - y(\hat{b}, c) = \int_{\hat{b}}^b F_1(t, t) dt.$$

Or, taking expectations,

$$E_c[y(b, c) - y(\hat{b}, c)] = \int_{\hat{b}}^b F_1(t, t) dt \leq \int_{\hat{b}}^b F_1(t, b) dt = F(b, b) - F(\hat{b}, b) \tag{29}$$

where the inequality follows from the assumption $F_{12} > 0$. Rearranging the above gives

$$F(b, b) - E_c[y(b, c)] \geq F(\hat{b}, b) - E_c[y(\hat{b}, c)] \tag{30}$$

which establishes IC for the buyer. IC for the seller goes through analogously.

To show IR it suffices to show $\tilde{U}(b_L) = 0$, since we know from (26) and the fact that $F_2(t, t)$ is non-decreasing in t that $\tilde{U}(b)$ is non decreasing. But by definition of y ,

$$y(b_L, c) = \int_{c_L}^c G_1(t, t) dt + K$$

where $K = F(b_L, b_L) - \int_C [1 - Q(t)] G_1(t, t) dt$ is a constant, independent of c . So

$$\begin{aligned} E_c[y(b_L, c)] &= \int_C \left[\int_{c_L}^c G_1(t, t) dt \right] q(c) dc + K = \int_C G_1(t, t) \left[\int_t^{c_H} q(c) dc \right] dt + K \\ &= \int_C [1 - Q(t)] G_1(t, t) dt + K = F(b_L, b_L) \end{aligned}$$

But by (25), $\tilde{U}(b_L) = F(b_L, b_L) - E_c[y(b_L, c)]$, so $\tilde{U}(b_L) = 0$. Finally, we need to prove IR for the seller. By assumption, $M(x) \geq 0$. By theorem 1, $M(x) = \tilde{U}(b_L) + \tilde{V}(c_H)$. From the definition of $y(b, c)$ we have $\tilde{U}(b_L) = 0$. Hence, $M(x) = \tilde{V}(c_H) \geq 0$. But $\tilde{V}(c)$ is non increasing. So $\tilde{V}(c) \geq 0$ for all $c \in C$.

(*Necessity*) From Theorem 1 together with IR, we have $M(x) \geq 0$. Next, we show the monotonicity conditions on F . By IC and (25), it follows that for any $b, \hat{b} \in B$, and $c, \hat{c} \in C$

$$\begin{aligned} F(b, b) - E_c[y(b, c)] &\geq F(\hat{b}, b) - E_c[y(\hat{b}, c)] \\ F(\hat{b}, \hat{b}) - E_c[y(\hat{b}, c)] &\geq F(b, \hat{b}) - E_c[y(b, c)]. \end{aligned}$$

These inequalities imply

$$F(b, b) - F(b, \hat{b}) \geq F(\hat{b}, b) - F(\hat{b}, \hat{b})$$

But if $b f_2(x, b) = f(x, b)$, then $b F_2(\hat{b}, b) = F(\hat{b}, b)$, and $f_{22}(x, b) = 0$ implies $F_{22}(\hat{b}, b) = 0$ for all b, \hat{b} , which implies that $F_2(\hat{b}, b) = F_2(\hat{b}, \hat{b})$. So ,

$$\begin{aligned} b F_2(b, b) - \hat{b} F_2(b, \hat{b}) &\geq b F_2(\hat{b}, b) - \hat{b} F_2(\hat{b}, \hat{b}) \Leftrightarrow \\ (b - \hat{b}) F_2(b, b) &\geq (b - \hat{b}) F_2(\hat{b}, b). \end{aligned}$$

which implies that F_2 , must be non decreasing in its first argument. I. e., $F_{21} \geq 0$. But then $\frac{\partial}{\partial t} F_2(t, t) = F_{21}(t, t) + F_{22}(t, t) = F_{21}(t, t) \geq 0$. A similar argument shows that $G_{21} = G_{12} \leq 0$ and $\frac{\partial}{\partial t} G_2(t, t) \leq 0$ ■

5.3 Proof of Theorem 4

Proof.

Consider two problems. The first problem is:

$$\max_{x \in \mathcal{X}} E[s(x; b, c)] \quad (31)$$

and the second problem is:

$$\begin{aligned} \max_{x \in \mathcal{X}} E[s(x; b, c)] \\ \text{s.t. } M(x) \geq 0 \end{aligned} \quad (32)$$

The first problem is solved by pointwise maximization, yielding $x^* = x^*(b, c)$. Recall from equation (5) that $s(x; b, c) = b\bar{f}(x) - c\bar{g}(x)$. Note that $s''(x; b, c) = b\bar{f}''(x) - c\bar{g}''(x) < 0$. So $s'(x; b, c)$ is decreasing in x . Hence, the solution to problem (31) must satisfy

$$s'(x^*; b, c) \begin{cases} = 0 & \text{if } 0 < x^* < x_H \\ \geq 0 & \text{if } x^* = x_H \\ \leq 0 & \text{if } x^* = 0 \end{cases} \quad (33)$$

For the second problem, we first show that any solution x^o of problem (32) satisfies $x^o(b, c) \leq x^*(b, c)$ a. e., with strict inequality on a set of positive measure. We then will show that any solution of problem (32) satisfies the additional monotonicity conditions on F and G of Theorem 3, and hence is indeed a maximally efficient mechanism.

The Lagrangean for problem (32) is:

$$L = E[s(x; b, c)] + \lambda M(x).$$

So

$$\begin{aligned} L &= E[b\bar{f}(x) - c\bar{g}(x)] \\ &\quad + \lambda E[(2b - 1)\bar{f}(x) - 2c\bar{g}(x)] \\ &= \int_C \int_B [(1 + 2\lambda)s(x(b, c); b, c) - \lambda\bar{f}(x(b, c))] db dc \end{aligned} \quad (34)$$

and $\lambda > 0$, because $M(x^*) < 0$. For any given λ , the solution to (32) must maximize the integrand of (34) almost everywhere. In other words, for almost all b, c , $x^o = x^o(b, c)$ must solve

$$\max_{x \in \mathbb{R}} (1 + 2\lambda)s(x; b, c) - \lambda\bar{f}(x) \quad (35)$$

Depending on the value of (b, c) this leads to one of three cases:

Case 1 In the region of (b, c) for which there is an interior solution ($0 < x^o < x_H$), of (35), a necessary for x to be a solution is,

$$(1 + 2\lambda)s'(x; b, c) - \lambda\bar{f}'(x) = 0 \quad (36)$$

Fix b and c , and assume that x satisfies $x^*(b, c) = x^* \leq x < x_H$. Then from equation (33), $s'(x^*; b, c) \leq 0$. And by monotonicity of s' , $s'(x; b, c) \leq s'(x^*; b, c) \leq 0$. Hence, the first term in (36) is negative or zero, and the second is negative, yielding a contradiction. So we must have $x^o < x^*$. Hence for almost all b, c , we have $x^o(b, c) < x^*(b, c)$.

Case 2 In the region of (b, c) for which equation (35) is satisfied for $x^o = x_H$, a necessary condition is for almost all b, c ,

$$(1 + 2\lambda)s'(x_H) - \lambda\bar{f}'(x_H) \geq 0$$

Then $s'(x_H) \geq \frac{\lambda}{1+2\lambda}\bar{f}'(x_H) \geq 0$. Since s' is strictly decreasing, this means $s'(x) > 0$ for all $x < x_H$. So $x^* = x^o$, implying that $x^o \leq x^*$.

Case 3 In the region of (b, c) for which equation (35) is satisfied for $x^o = 0$, then we automatically have $x^o \leq x^*$, with strict inequality whenever $x^* \neq 0$.

We have shown that in all three cases we have $x^o(b, c) \leq x^*(b, c)$ for almost all b, c . Now by assumption, there is a set of positive measure where $0 < x^*(b, c) < x_H$. Denote this set by D . Then D cannot intersect with case 2, since in that case we must have $x^* = x^o = x_H$. So D must be a subset of cases 1 and 3. But in both of these cases, we showed that $x^o(b, c) < x^*(b, c)$ for almost all b, c . Hence, there must be a set of positive measure where $x^o(b, c) < x^*(b, c)$.

We next show that any solution of problem (31) must also satisfy the monotonicity conditions on \bar{F} and \bar{G} . To see this, we show that $\frac{\partial}{\partial b}x(b, c) \geq 0$ and $\frac{\partial}{\partial c}x(b, c) \leq 0$. Consider any interior solution. Then we can rewrite equation (36) as

$$b = c \frac{\bar{g}'(x)}{f'(x)} + K,$$

where $K = \frac{\lambda}{1+2\lambda}$. Taking derivatives with respect to b , we get

$$1 = c \left(\frac{\bar{f}'(x)\bar{g}''(x) - \bar{g}'(x)\bar{f}''(x)}{[\bar{f}'(x)]^2} \right) \frac{\partial x}{\partial b},$$

Since the term in parenthesis is positive, and c is positive, it follows that $\frac{\partial x}{\partial b}$ must also be positive. Similarly, taking derivatives with respect to c , we get

$$0 = \frac{\bar{g}'(x)}{f'(x)} + c \left(\frac{\bar{f}'(x)\bar{g}''(x) - \bar{g}'(x)\bar{f}''(x)}{[\bar{f}'(x)]^2} \right) \frac{\partial x}{\partial c}$$

Since the first term is positive, and the term in parenthesis is positive, $\frac{\partial x}{\partial c}$ must be less than zero. In the regions where x is not interior, it must be constant, hence, we get $\frac{\partial}{\partial b}x(b, c) \geq 0$ and $\frac{\partial}{\partial c}x(b, c) \leq 0$ trivially in those regions. But then $\bar{F}'(b) = \frac{\partial}{\partial b}E_c[\bar{f}(x(b, c))] = E_c[\frac{\partial}{\partial b}\bar{f}(x(b, c))] \geq 0$. Similarly, $\bar{G}'(c) \leq 0$. ■

5.4 Proof of Proposition 5

5.4.1 Part A

Proof. When $B = C = [0, 1]$, we substitute into (15) to get

$$M(x^*) = \frac{1}{3} [\ln(2) - 1] < 0$$

But then, by Corollary 1, it follows that there is no efficient solution. ■

5.4.2 Part B

Proof. It follows from Corollary 1 that if $(x, y) \in \mathcal{X} \times \mathcal{X}$ satisfies IC and IR, then we must have $M(x) \geq 0$. Hence, to show that the expected x is less than the efficient x , it suffices to show that if the expected x is greater than or equal to the efficient x , then we

can't obtain $M(x) \geq 0$. So let \mathcal{X} be the set of integrable functions on $T = [0, 1] \times [0, 1]$.

This is the set of L^1 functions on T . Thus, we consider the problem:

$$\begin{aligned}
& \max_{x \in \mathcal{X}} && M(x) \\
& \text{subject to} && \int_0^1 \int_0^1 x(b, c) db dc \geq \int_0^1 \int_0^1 \frac{b}{b+c} db dc = \frac{1}{2} \\
& && x(b, c) \geq 0 \\
& && x(b, c) \leq 1
\end{aligned} \tag{37}$$

We will show that the value of the objective function at the solution to this optimization problem is negative.

Necessary conditions for a solution to this problem can be obtained from the generalized Kuhn Tucker theorem (see e. g., Luenberger [1969], Theorem 1, p. 249). The Lagrangean for the above problem is

$$\begin{aligned}
L(x, \lambda, \gamma_0, \gamma_1) &= \int_0^1 \int_0^1 [(1 - 2b - 2c)x^2(b, c) + (4b - 2)x(b, c)] db dc \\
&+ \lambda \left[\int_0^1 \int_0^1 x(b, c) db dc - \frac{1}{2} \right] + \left[\int_0^1 \int_0^1 x(b, c) \gamma_0(b, c) db dc \right] \\
&+ \left[\int_0^1 \int_0^1 (1 - x(b, c)) \gamma_1(b, c) db dc \right]
\end{aligned} \tag{38}$$

where $\lambda \in \Re$ is non negative, and γ_0, γ_1 are non negative L^∞ functions on T . Here, we use the duality of L^1 and L^∞ (see Aliprantis and Border, Theorem 10.28, p. 354).

A necessary condition for x to be a solution is that it be a stationary point of the Lagrangean. In other words, we must have

$$\left. \frac{\partial}{\partial \alpha} L(x + \alpha h, \lambda, \gamma_0, \gamma_1) \right|_{\alpha=0} = 0$$

for all $h \in L^\infty$. But

$$\begin{aligned}
\left. \frac{\partial}{\partial \alpha} L(x + \alpha h, \lambda, \gamma_0, \gamma_1) \right|_{\alpha=0} &= \int_0^1 \int_0^1 [2(1 - 2b - 2c)x(b, c)h(b, c) + (4b - 2)h(b, c)] db dc \\
&+ \lambda \int_0^1 \int_0^1 h(b, c) db dc + \int_0^1 \int_0^1 h(b, c) \gamma_0(b, c) db dc \\
&- \int_0^1 \int_0^1 h(b, c) \gamma_1(b, c) db dc
\end{aligned} \tag{39}$$

But if $h = \chi_A$, where $A = [0, b] \times [0, c]$, then the above reduces to:

$$\begin{aligned} 0 &= \int_0^c \int_0^b [2(1 - 2b - 2c)x(b, c) + (4b - 2)] db dc \\ &\quad + \lambda bc + \int_0^c \int_0^b \gamma_0(b, c) db dc - \int_0^c \int_0^b \gamma_1(b, c) db dc \end{aligned} \quad (40)$$

for all $b, c \in [0, 1]$. Now taking derivatives of the above with respect to b and c , and appealing to the Fundamental Theorem of Calculus, we get necessary conditions for a solution are

$$\begin{aligned} &2(1 - 2b - 2c)x(b, c) + 4b - 2 + \lambda + \gamma_0(b, c) - \gamma_1(b, c) \\ &= 4(b - \frac{1}{2} + \frac{\lambda}{4})(1 - x(b, c)) - 4(c - \frac{\lambda}{4})x(b, c) + \gamma_0(b, c) - \gamma_1(b, c) \end{aligned} \quad (41)$$

$$= 0 \quad (42)$$

Further, the constraints must be satisfied:

$$\lambda \left(\int_0^1 \int_0^1 x(b, c) db dc - \frac{1}{2} \right) = 0 \quad (43)$$

$$\int_0^1 \int_0^1 x(b, c) \gamma_0(b, c) db dc = 0 \quad (44)$$

$$\int_0^1 \int_0^1 (1 - x(b, c)) \gamma_1(b, c) db dc = 0 \quad (45)$$

From (44) and (45), it follows that if $0 < x(b, c) < 1$, then $\gamma_0(b, c) = \gamma_1(b, c) = 0$. From the second order conditions, it follows that any solution satisfying (41) is a maximum only if $b + c \geq \frac{1}{2}$. Using these facts, we conclude that any solution to (37) must satisfy:

$$x(b, c) = \begin{cases} 0 & \text{if } b - c < \frac{1}{2}(1 - \lambda), b < \frac{1}{2} - \frac{\lambda}{4} \\ 1 & \text{if } b - c \geq \frac{1}{2}(1 - \lambda), c < \frac{\lambda}{4} \\ \frac{b - \frac{1}{2} + \frac{\lambda}{4}}{b + c - \frac{1}{2}} & \text{otherwise} \end{cases} \quad (46)$$

Now, consider the special case $\lambda = 1$. In this case,

$$x(b, c; \lambda) = \begin{cases} 0 & \text{if } b < c, \text{ and } b < \frac{1}{4} \\ 1 & \text{if } b > c, \text{ and } c < \frac{1}{4} \\ \frac{b + \frac{1}{4}}{b + c - \frac{1}{2}} & \text{otherwise} \end{cases} \quad (47)$$

and $x(b, c; 1) + x(c, b; 1) = 1$, which means that $\int_0^1 \int_0^1 x(b, c) db dc = \frac{1}{2}$. So $x(b, c; 1)$ is a solution to (43).

It is easy to check that this solution is unique. From (46), note that for $x(b, c; \lambda) = 0$ and $x(b, c; \lambda) = 1$, $x(b, c; \lambda)$ is non-decreasing in λ ; and when $\lambda = 1$, $0 < x(b, c; \lambda) < 1$ for a set of positive measure. Thus, for all $\lambda \geq 0$, $\int_0^1 \int_0^1 x(b, c) db dc$ is non-decreasing in λ , and is strictly increasing in λ for λ in a neighborhood of 1. Thus, $x(b, c; 1)$ is the only solution to (37).

But now for $\lambda = 1$, the value of the objective function is:

$$\begin{aligned} M(x) &= \int_0^1 \int_0^1 [(1 - 2b - 2c)x(b, c)^2 + (4b - 2)x(b, c)] db dc \\ &= \int_{\frac{1}{4}}^1 \int_{\frac{1}{4}}^1 \frac{2(b - \frac{1}{4})(b - \frac{3}{4})}{b + c - \frac{1}{2}} db dc + \int_0^{\frac{1}{4}} \int_c^1 (2b - 2c - 1) db dc \end{aligned} \quad (48)$$

The first term evaluates to

$$\begin{aligned} & \int_{\frac{1}{4}}^1 2(b - \frac{1}{4})(b - \frac{3}{4}) \ln(2b + 2c - 1) \Big|_{\frac{1}{4}}^1 db \\ &= \int_{\frac{1}{4}}^1 2(b - \frac{1}{4})(b - \frac{3}{4}) \ln\left(\frac{2b + 1}{2b - \frac{1}{2}}\right) db \\ &= \left[\frac{1}{3}(2b + 1)\left[(b - 1)^2 + \frac{9}{16}\right] \ln(2b + 1) \right. \\ & \quad \left. - \frac{2}{3}\left(b - \frac{1}{4}\right)^2(b - 1) \ln\left(2b - \frac{1}{2}\right) + \frac{b^2}{4} - \frac{7b}{8} \right] \Big|_{\frac{1}{4}}^1 \\ &= \frac{9}{16} \ln(2) - \frac{27}{64} = \left(\frac{3}{4}\right)^2 (\ln(2) - \frac{3}{4}) < 0 \end{aligned} \quad (49)$$

and the second term of (48) evaluates to $-\frac{5}{192}$. Hence, we have

$$M(x) = \frac{9}{16} \ln(2) - \frac{27}{64} - \frac{5}{192} < 0 \quad (50)$$

which completes the proof. ■