Efficient Mechanisms for Level-*k* Bilateral Trading

Vincent P. Crawford[1]

First version 18 April 2015; this revision 14 March 2016

Abstract: I revisit Roger Myerson and Mark Satterthwaite's (1983; "MS") analysis of mechanism design for bilateral trading, replacing equilibrium with a nonequilibrium "level-*k*" model that predicts initial responses to games, and focusing on direct mechanisms. The revelation principle fails for level-*k* models. However, if only level-*k*-incentive-compatible mechanisms are feasible, MS's result that no incentive-compatible mechanism can assure ex post efficiency generalizes. If traders' levels are observable, MS's characterization of incentive-efficient mechanisms generalizes, with one novel feature. If they're unobservable, only posted-price mechanisms are level-*k*-incentive-compatible and -incentive-efficient. If non-level-*k*-incentive-compatible direct mechanisms are feasible, level-*k*-incentive-efficient mechanisms may differ more extensively from equilibrium-incentive-efficient mechanisms.

Keywords: mechanism design, incentive-efficient bilateral trading, revelation principle, behavioral game theory, level-*k* thinking

---

[1] Department of Economics, University of Oxford, Manor Road Building, Oxford OX1 3UQ, UK; All Souls College, Oxford OX1 4AL, UK; and Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA; e-mail: vincent.crawford@economics.ox.ac.uk. This is an extensively revised version of the first version, dated 14 April 2015. It corrects significant errors in Theorem 2 and some of the later results as indicated below. Special thanks are due to Rene Saran for pointing out those errors and aiding my efforts to correct them. I also owe thanks for valuable advice to Johannes Abeler, Douglas Bernheim, Colin Camerer, Yan Chen, Miguel Costa-Gomes, Jacques Crémer, Avinash Dixit, Kfir Eliaz, Tore Ellingsen, Xavier Gabaix, Olga Gorelkina, Nagore Iriberri, Paul Klemperer, Margaret Meyer, Paul Milgrom, Roger Myerson, Philip Neary, Zvika Neeman, Robert Östling, Ady Pauzner, Phil Reny, Adam Sanjurjo, Joel Sobel, Alex Teytelboym, and four referees; and to Rustu Duran for outstanding research assistance. The University of Oxford, All Souls College, and the University of California, San Diego provided research support. The research leading to these results received primary funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 339179. The contents reflect only the author's views and not the views of the ERC or the European Commission, and the European Union is not liable for any use that may be made of the information contained therein.

This paper revisits Myerson and Satterthwaite's (1983; "MS") classic analysis of mechanism design for bilateral trading with independent private values. I focus on direct mechanisms and replace MS's assumption that traders will play the desired equilibrium in any game the choice of mechanism creates with the assumption that traders will follow a structural nonequilibrium model based on level-*k* thinking. Otherwise I maintain standard behavioral assumptions.

Equilibrium-based analyses of design have enjoyed tremendous success; and both theory and experiments support the assumption that players in a game who have enough experience with analogous games will have learned to play an equilibrium. Why, then, study nonequilibrium design? Design creates new games, which may lack the clear precedents required for learning; yet a design may still need to work the first time. Further, even if learning is possible, design may create games too complex for convergence to equilibrium to be behaviorally plausible.

In theory, assuming equilibrium can still be justified via epistemic arguments (Robert Aumann and Adam Brandenburger 1995). But in experiments that study initial responses to games, subjects' thinking seldom follows the fixed-point or iterated-dominance reasoning that equilibrium usually requires.[2] Instead their thinking often favors level-*k* decision rules, which anchor beliefs in a naive model of others' initial responses called *L0*, and then adjust them via a small number (k) of iterated best responses: *L1* best responds to *L0*, *L2* to *L1*, and so on. Estimates suggest that the frequency of *L0* is zero or small; and that subjects' levels are heterogeneous, concentrated on levels *L1*, *L2*, and *L3*.

Lk for $k > 0$ is rational in the decision-theoretic sense, with an accurate model of the game. It departs from equilibrium only in basing beliefs on an oversimplified model of others' decisions. *Lk*'s decisions also respect *k*-rationalizability (Bernheim 1984), so that *Lk* mimics equilibrium decisions in two-person games that are dominance-solvable in *k* rounds, but can deviate systematically in other games.[3] Importantly, the resulting model not only predicts that deviations from equilibrium will sometimes occur, but also which kinds of game evoke them and what

[2] Crawford, Costa-Gomes, and Iriberri (2013) survey the experimental literature on strategic thinking. Some researchers argue that using an incentive-compatible mechanism and announcing that truth-telling is an equilibrium avoids the complexity of equilibrium thinking, but people may wish to check such claims via their own thinking. Eric Maskin (2011) argues that "the theoretical and practical drawbacks of Nash equilibrium as a solution concept are far less troublesome in problems of mechanism design", because the game can often be chosen to ensure that equilibrium is unique, or even that it is dominance-solvable. But the experiments suggest that neither of those features assures equilibrium initial responses in games of the kind used in analyses of implementation (Elena Katok, Martin Sefton, and Abdullah Yavas 2002; Chen and John Ledyard 2008).

[3] In Camerer, Teck-Hua Ho, and Juin-Kuan Chong's (2004) closely related "cognitive hierarchy" model, *Lk* best responds to an estimated mixture of all lower levels. A cognitive hierarchy *Lk* need not always respect *k*-rationalizability when $k > 1$.

forms they will take. It also replaces *k*-rationalizability's set-valued predictions with a specific selection that permits an analysis with precision close to that of an equilibrium analysis.[4]

I focus on the level-*k* model because it is well supported by evidence, and for concreteness. But as indicated below, most of my results also hold for alternative nonequilibrium models that make unique, well-behaved predictions that can be viewed as best responses to some beliefs.

A structural nonequilibrium analysis of mechanism design has several potential benefits. It can clarify the role of equilibrium assumptions in analyses like MS's. It can identify settings in which equilibrium conclusions are robust to empirically likely deviations from equilibrium; and other settings where mechanisms that are optimal if equilibrium is assumed, perform worse in practice than others whose performance is more robust to deviations. Finally, a nonequilibrium analysis might reduce the sensitivity of incentive-efficient mechanisms to distributional and knowledge assumptions that real institutions seldom respond to (Robert Wilson 1987).

Section I reviews the positive starting point for MS's analysis, Kalyan Chatterjee and William Samuelson's (1983; "CS") equilibrium analysis of bilateral trading with independent private values via double auction. CS characterized the equilibria of double auctions when traders have well-behaved value densities with overlapping supports. When the densities are uniform, CS gave a closed-form solution for an equilibrium in which traders' bids are linear in their values, in which they shade their bids so that trade occurs if and only if the buyer's value is sufficiently larger than the seller's. In this and other equilibria, with positive probability some beneficial trades do not occur and trading is ex post Pareto-inefficient.

MS asked whether the ex post inefficiency that CS noted is an avoidable flaw of the double auction or a general property of any incentive-feasible trading mechanism with private values. Section II reviews MS's analysis of mechanism design. They argued, via the revelation principle (MS, pp. 267-268), that any equilibrium of any mechanism can be viewed as the truthful equilibrium of some direct-revelation mechanism. Thus, assuming equilibrium, there is no loss of generality in restricting attention to direct mechanisms that are incentive-compatible in the sense that truthful reporting of values is an equilibrium. MS characterized incentive-efficient trading mechanisms for well-behaved value densities with overlapping supports, showing that in

[4] Until recently the alternatives to assuming equilibrium were limited to quantal response equilibrium and rationalizability or *k*-rationalizability. To my knowledge quantal response equilibrium has not been applied to design, perhaps because its predictions must be solved for numerically and are sensitive to its error structure. Rationalizability and *k*-rationalizability have been applied to design as noted below, but the level-*k* model's selection allows an analysis that yields additional insight.

equilibrium, the ex post inefficiency of CS's double auction cannot be avoided by any mechanism. MS also showed that with uniform densities and symmetric surplus-sharing, CS's linear double-auction equilibrium, or equivalently (via the revelation principle) the incentive-compatible direct mechanism that mimics its outcomes, yields an incentive-efficient outcome.

Section III defines a level-*k* model for incomplete-information games. I focus on games such as those created by direct mechanisms, in which players' decisions can be viewed as estimates of their values. The simplicity of such games is advantageous in applications; and for them there is clear evidence to guide the specification of a level-*k* model. With complete information, *L0*'s decisions are usually taken to be uniformly random over the feasible decisions. With incomplete information I take *L0*'s decisions to be uniform over the feasible decisions and *independent of its own private value*. As usual, I define *L1*, *L2*, etc., via iterated best responses. This generalized level-*k* model gives a reliable account of how people's thinking deviates from equilibrium, and also of their "informational naiveté"—the imperfect attention to how others' decisions depend on their private information commonly observed in phenomena like the winner's curse.[5]

To set the stage for Section V's analysis of level-*k* design, Section IV revisits CS's equilibrium analysis of the double auction using Section III's level-*k* model. In Section IV I focus on *L1*s or *L2*s, which are prevalent empirically and illustrate my main points; and also on the leading case of uniform value densities. In that case *L1*s' beliefs about others' bids or asks are too optimistic, relative to equilibrium, which makes them ask or bid too aggressively. That aggressiveness drives *L1*s' expected total surplus far below its level in CS's linear equilibrium. By contrast, *L2*s' beliefs are too pessimistic, which makes them bid or ask too *un*aggressively, raising *L2*s' expected total surplus above its level in CS's linear equilibrium.

Section V begins my analysis of level-*k* design, which focuses on direct mechanisms throughout. I use "incentive-compatible" for mechanisms that make it optimal for traders to report their values truthfully, given their beliefs; and "level-*k*-incentive-compatibility" etc. for the level-*k* analogues of the standard notions, which I call "equilibrium-incentive-compatibility" etc. In Section V I assume level-*k*-incentive-compatibility is required, which entails loss of generality as explained below.[6] Theorem 1 records the fact that with uniform value densities, MS's equilibrium-incentive-efficient mechanism is efficient in the set of level-*k*-incentive-

[5] Imperfect attention to how others' decisions depend on their private information is highly relevant here, even though the winner's curse is not relevant for trading with independent private values.
[6] Section VI relaxes the requirement of level-k-incentive-compatibility.

compatible mechanisms (then independent of k) for any population of level-k traders with $k > 0$, whether or not the designer knows or can observe individual traders' levels. Thus in this case MS's characterization of incentive-efficient mechanisms is fully robust to level-k thinking.

With general well-behaved (or uniform) value densities, Corollary 1 extends MS's result that no equilibrium-incentive-compatible mechanism can assure ex post efficiency with probability one to a wide class of level-k models. MS's result does not depend on equilibrium or, as the proof shows, even the special structure of level-k models: It holds for any model that makes unique, well-behaved predictions that can be viewed as best responses to some beliefs.

Theorem 1 shows only that with uniform value densities, MS's equilibrium-incentive-efficient mechanism is efficient in the set of *Lk*-incentive-compatible mechanisms if implemented in its *Lk*-incentive-compatible direct form: Section IV's analysis shows that the double auction yields lower surplus for *L1*s than MS's equilibrium-incentive-efficient mechanism, thus violating the revelation principle; and that the double auction yields higher surplus for *L2*s, violating the revelation principle in a different way. Replacing the double auction with MS's mechanism rectifies *L1*s' beliefs, eliminating their aggressiveness and yielding higher surplus. But the double auction preserves *L2*s' beneficial pessimism, yielding higher surplus than in MS's mechanism, or any mechanism that is efficient in the set of *L2*-incentive-compatible mechanisms.[7] The revelation principle fails for level-k traders because of Crawford, Tamar Kugler, Zvika Neeman, and Ady Pauzner's (2009) "level-k menu effects", whereby the choice of mechanism affects the correctness of level-k beliefs, which are anchored on an *L0* whose effect is not eliminated by equilibrium thinking.

Because the revelation principle fails for level-k traders, it matters whether level-k-incentive-compatibility is truly required or can be relaxed to allow non-incentive-compatible direct mechanisms like the double auction. Here I do not ask whether incentive-compatibility is required, which is an empirical question.[8] Instead Section V continues to consider the design problem when level-k-incentive-compatibility is required, and Section VI considers relaxing it.[9]

[7] The latter result shows that nonequilibrium design involves more than implementing equilibrium outcomes under weaker behavioral assumptions, as is sometimes assumed in the literature on robust mechanism design. In another example of this, Crawford et al. (2009) showed that, although a second-price auction may seem preferable to a first-price auction because it yields equilibrium outcomes for level-k as well as equilibrium bidders, revenue-equivalence fails for level-k bidders, and a first-price auction, especially if *L1*s are prevalent, tends to yield higher expected revenue than a second-price auction.

[8] Many analysts argue that incentive-compatibility is essential in applications (Milgrom, Larry Ausubel, Jon Levin, and Ilya Segal 2012 for auctions; Atila Abdulkadiroglu and Tayfun Sönmez 2003 for school choice), though mostly in equilibrium analyses where there is no theoretical gain from relaxing it. Other analysts are willing to consider non-incentive-compatible

When level-*k*-incentive-compatibility is required, the results for general value densities depend on whether the designer knows or can observe individual traders' levels, in which case s/he can enforce a mechanism tailored to the buyer's and seller's levels. Theorems 2-3 show that with observable levels, MS's characterization of incentive-efficient mechanisms fully generalizes to the level-*k* model. Thus the design features that foster equilibrium-incentive-efficiency also foster efficiency in the set of level-*k*-incentive-compatible mechanisms. The level-*k* analysis adds a novel non-equilibrium feature, "tacit exploitation of predictably incorrect beliefs" ("TEPIB"), which skews the optimal mechanism in ways that yield more surplus than with correct beliefs.[10]

Theorem 4 turns to the case where traders' levels are heterogeneous and the designer cannot observe them, and so must screen traders' levels along with their values. Level-*k*-incentive-compatibility then compels the use of a random posted-price mechanism (Kathleen Hagerty and William Rogerson 1987; Jernej Čopič and Clara Ponsatí 2008, 2015); and any mechanism that is efficient in the set of level-*k*-incentive-compatible mechanisms is equivalent to a deterministic posted-price mechanism. This neutralizes TEPIB, while coming closer to real institutions.

Together, Theorems 2-4 show that if incentive-compatibility is required, an analysis like MS's can dispense with the equilibrium assumption, as long as individual traders follow well-behaved decision rules that can be viewed as best responding to some beliefs. Put another way, if incentive-compatibility is required, equilibrium's power to simplify such analyses comes mainly from its implication that the designer knows traders' decision rules.

If incentive-compatibility is not required, equilibrium simplifies analyses in a different way, implying via the revelation principle that relaxing incentive-compatibility is irrelevant. Section VI takes up this aspect of equilibrium analysis by studying level-*k* design while relaxing level-*k*-incentive-compatibility, continuing to focus on direct mechanisms. Here one can still define a general class of direct mechanisms; but a mechanism's incentive effects can no longer tractably be captured via constraints, but must be modeled directly via level-*k* traders' responses to it.

If traders' levels are observable by the designer, the value densities are uniform, and menus of bids or asks restricted by reserve prices make level-*k* traders anchor on the restricted menu instead of on [0, 1], for *L1*s reserve prices chosen optimally by the designer allow the non-level-

mechanisms such as first-price sealed-bid auctions (Myerson 1981) or the Boston Mechanism (Aytek Erdil and Haluk Ergin 2008; Abdulkadiroglu, Yeon-Koo Che, and Yosuke Yasuda 2011).

[9] "Relaxing it" means allowing direct mechanisms that create incentives to lie, but with traders still assumed to best respond.

[10] "Predictably incorrect" because the level-*k* model predicts traders' deviations from equilibrium; "exploitation" in the benign sense of using traders' nonequilibrium responses for their own benefit; and "tacit" in that traders are not actively misled.

k-incentive-compatible double auction to rectify their beliefs and yield the outcome of MS's equilibrium-incentive-efficient mechanism. For *L2*s a double auction without reserve prices improves upon any mechanism that is efficient in the set of *L2*-incentive-efficient mechanisms or MS's equilibrium-incentive-efficient mechanism; but reserve prices can do no better.

If traders' levels are heterogeneous and unobservable, but the designer knows the population frequency of one, known level is high, it can be beneficial to design for that level while ignoring incentive constraints for other levels. More generally, relaxing level-*k*-incentive-compatibility can yield level-*k*-incentive-efficient mechanisms that differ qualitatively as well as quantitatively from equilibrium-incentive-efficient mechanisms, with substantial gains in incentive-efficiency.

In addition to CS's and MS's analyses, this paper builds on Crawford and Iriberri's (2007) positive level-*k* analysis of sealed-bid auctions and Crawford et al.'s (2009) level-*k* analysis of optimal independent-private-value auctions, which builds in turn on Myerson's (1981) classic equilibrium analysis. In other related work, Terri Kneeland (2013), Olga Gorelkina (2015), and Geoffroy de Clippel, Saran, and Roberto Serrano (2015) use level-*k* models to study design in settings including bilateral trading. Saran (2011a) studies MS's design problem when some traders report truthfully without regard to incentives. Hagerty and Rogerson (1987), Bulow and Roberts (1989, giving up ex post budget balance), and Čopič and Ponsatí (2008, 2015) study dominant-strategy or distribution-free implementation in MS's trading environment. Paul Healy (2006) studies design of public goods mechanisms when agents must learn an equilibrium. In more abstract settings, Dilip Mookherjee and Stefan Reichelstein (1992) study dominant-strategy implementation; Hitoshi Matsushima (2007, 2008) studies implementation via finitely iterated dominance; and Dirk Bergemann, Stephen Morris, and Olivier Tercieux (2011) study implementation in rationalizable strategies. Finally, Kobi Glazer and Ariel Rubinstein (1998), Neeman (2003), Kfir Eliaz and Ran Spiegler (2006, 2007, 2008), and Alexander Wolitzky (2014) study design when the "behavioral" aspect concerns individual decisions or judgment.

I. EQUILIBRIUM BILATERAL TRADING VIA DOUBLE AUCTION

Following CS and MS, I consider bilateral trading between a potential seller and buyer of an indivisible object, in exchange for an amount of money to be determined. The traders' von Neumann-Morgenstern utility functions are quasilinear in money, so they are risk-neutral and have well-defined money values for the object. Denote the buyer's value V and the seller's C (for "cost"; but I sometimes use "value" generically for C as well as V). V and C are independently distributed, with probability densities $f(V)$ and $g(C)$ that are strictly positive on their supports, and probability distribution functions $F(V)$ and $G(C)$. CS and MS allowed traders' value distributions to have any bounded overlapping supports, but for simplicity and with no important loss of generality, I take their supports to be identical and normalize them to [0, 1].

CS study a double auction, in which traders make simultaneous money offers. If the buyer's offer b (for "bid") exceeds the seller's offer a ("ask"), they exchange the object for a price that is a weighted average of a and b. CS allowed any weights from 0 to 1, but as in MS's analysis I focus on the symmetric case with weights ½. Then, if $b \geq a$, the buyer acquires the object at price $(a + b)/2$, the seller's utility is $(a + b)/2$, and the buyer's is $V - (a + b)/2$. If $b < a$, the seller retains the object, no money changes hands, the seller's utility is C, and the buyer's is 0.

As CS noted, this game has many Bayesian equilibria. I follow them and the subsequent literature in focusing on equilibria in which trade occurs with positive probability, and traders' strategies are bounded above and below, strictly increasing, and (except possibly at the boundaries) differentiable. Denote the buyer's bidding strategy $b(V)$ and the seller's asking strategy $a(C)$. An equilibrium buyer's bid $b_*(V)$ must maximize, over $b \in [0, 1]$

$$\int_0^b \left(V - \left[\frac{a+b}{2}\right]\right) g(a_*^{-1}(a))da + \int_b^1 0\, da,$$

where $g(a_*^{-1}(a))$ is the density of an equilibrium seller's ask $a_*(C)$ induced by the seller's value density $g(C)$. Similarly, an equilibrium seller's ask $a_*(C)$ must maximize, over $a \in [0, 1]$

$$\int_a^1 \left[\frac{a+b}{2}\right] f(b_*^{-1}(b))db + \int_0^a C\, f(b_*^{-1}(b))db,$$

where $f(b_*^{-1}(b))$ is the density of an equilibrium buyer's bid $b_*(V)$ given the value density $f(V)$.

In the leading case where traders' value densities $f(V)$ and $g(C)$ are uniform, CS gave a closed-form solution for a linear equilibrium, which was also important in MS's analysis. Given my normalization of the supports of $f(V)$ and $g(C)$ to [0, 1], in this equilibrium $b_*(V) = 2V/3 + 1/12$ unless $V < 1/4$, in which case $b_*(V)$ can be anything that does not lead to trade; and $a_*(C) = 2C/3 + 1/4$ unless $C > 3/4$, when $a_*(C)$ can be anything that does not lead to trade.

With those strategies, trade takes place if and only if $2V/3 + 1/12 \geq 2C/3 + 1/4$, or $V \geq C + 1/4$, at price $(V + C)/3 + 1/6$. Thus with positive probability the outcome is ex post inefficient. Even so, MS showed that in this case the double auction implements an ex ante incentive-efficient outcome: assuming equilibrium, no mechanism can ex ante Pareto-dominate the linear equilibrium of the double auction (MS's p. 277; my Section II.C). That linear equilibrium yields ex ante probability of trade $9/32 \approx 28\%$ and expected surplus $9/64 \approx 0.14$, less than the maximum ex post individually rational probability of trade of 50% and expected total surplus $1/6 \approx 0.17$.

II. EQUILIBRIUM MECHANISM DESIGN FOR BILATERAL TRADING

Assuming Bayesian equilibrium, MS characterized ex ante incentive-efficient mechanisms in CS's trading environment, allowing any feasible mechanism and taking into account the need to ensure interim individual rationality. I now review MS's analysis, using my notation.

A. *The revelation principle*

In a *direct* mechanism traders make simultaneous reports of their values, which I denote v and c to distinguish them from traders' true values V and C, and those reports then determine the outcome. MS's assumption that traders will play the desired equilibrium in any game the designer's choice of mechanism creates allows an important simplification of their analysis via the revelation principle. Because the revelation principle must be reconsidered in the level-k analysis, I quote MS's (pp. 267-268) equilibrium-based argument for that simplification here:

> We can, without any loss of generality, restrict our attention to incentive-compatible direct mechanisms. This is because, for any Bayesian equilibrium of any bargaining game, there is an equivalent incentive-compatible direct mechanism that always yields the same outcomes (when the individuals play the honest equilibrium)….[w]e can construct [such a] mechanism by first asking the buyer and seller each to confidentially report his valuation, then computing what each would have done in the given equilibrium strategies with these valuations, and then implementing the outcome (transfer of money and object) as in the given game for this computed behavior. If either individual had any incentive to lie to us in this direct mechanism, then he would have had an incentive to lie to himself in the original game, which is a contradiction of the premise that he was in equilibrium in the original game.

B. *Equilibrium-incentive-compatible direct trading mechanisms*

When traders are risk-neutral, the payoff-relevant outcomes of a direct mechanism are completely described by two outcome functions, $p(\cdot, \cdot)$ and $x(\cdot, \cdot)$, where if the buyer and seller report values v and c, then $p(v, c)$ is the probability the object is transferred from seller to buyer and $x(v, c)$ is the expected monetary payment from buyer to seller.[11] For a direct mechanism

[11] Thus MS assume the mechanism satisfies ex post (expected) budget balance, as I will assume in Section V's level-k analysis.

$p(\cdot,\cdot)$ and $x(\cdot,\cdot)$, define the buyer's and seller's expected monetary payments, probabilities of trade, and expected gains from trade as functions of their value reports v and c and true values V and C (with hats denoting variables of integration throughout whenever it is helpful for clarity):

$$X_B(v) = \int_0^1 x(v,\hat{c})g(\hat{c})d\hat{c}, \qquad X_S(c) = \int_0^1 x(\hat{v},c)f(\hat{v})d\hat{v},$$

(2.1) $$P_B(v) = \int_0^1 p(v,\hat{c})g(\hat{c})d\hat{c}, \qquad P_S(c) = \int_0^1 p(\hat{v},c)f(\hat{v})d\hat{v},$$

$$U_B(V,v) = VP_B(v) - X_B(v), \qquad U_S(C,c) = X_S(c) - CP_S(c).$$

Although the outcome functions take only on traders' reported values as arguments, traders' expected utilities also depend on their true values. Thus the mechanism $p(\cdot,\cdot)$, $x(\cdot,\cdot)$ (with the qualification "direct" omitted from now on) is *incentive-compatible* if and only if truthful reporting is an equilibrium; that is, if for every V, v, C, and c in [0, 1],

(2.2) $U_B(V,V) \geq U_B(V,v) = VP_B(v) - X_B(v)$ and $U_S(C,C) \geq U_S(C,c) = X_S(c) - CP_S(c)$.

Given incentive-compatibility, $p(\cdot,\cdot)$, $x(\cdot,\cdot)$ is *interim individually rational* if and only if for every V and C in [0, 1],

(2.3) $$U_B(V,V) \geq 0 \text{ and } U_S(C,C) \geq 0.$$

MS's Theorem 1. *For any incentive-compatible mechanism,*

(2.4) $$U_B(0,0) + U_S(1,1) = \min_{V\in[0,1]} U_B(V,V) + \min_{C\in[0,1]} U_S(C,C)$$

$$= \int_0^1\int_0^1 \left(\left[V - \frac{1-F(V)}{f(V)}\right] - \left[C + \frac{G(C)}{g(C)}\right]\right)p(V,C)g(C)f(V)dCdV.$$

Furthermore, if $p(\cdot,\cdot)$ is any function mapping [0, 1]×[0, 1] into [0, 1], then there exists a function $x(\cdot,\cdot)$ such that (p, x) is incentive-compatible and interim individually rational if and only if $P_B(\cdot)$ is weakly increasing, $P_S(\cdot)$ is weakly decreasing, and

(2.5) $$0 \leq \int_0^1\int_0^1 \left(\left[V - \frac{1-F(V)}{f(V)}\right] - \left[C + \frac{G(C)}{g(C)}\right]\right)p(V,C)g(C)f(V)dCdV.$$

Proof. MS (pp. 269-270) showed that (2.2) implies that $P_B(\cdot)$ is weakly increasing and $P_S(\cdot)$ is weakly decreasing and yields necessary and sufficient conditions for incentive-compatibility:

(2.6) $U_B(V,V) = U_B(0,0) + \int_0^V P_B(\hat{v})d\hat{v}$ and $U_S(C,C) = U_S(1,1) + \int_C^1 P_S(\hat{c})d\hat{c}$ for all V and C.

(2.6) implies that $U_B(V,V)$ is weakly increasing and $U_S(C,C)$ is weakly decreasing, so that $U_B(0,0) \geq 0$ and $U_S(1,1) \geq 0$ suffice for interim individual rationality as in (2.3). MS (p. 270) next showed (via algebra that is a special case of that in Section V.C's proof of Theorem 2) that

(2.7) $$\int_0^1 \int_0^1 (V-C)\, p(V,C) g(C) f(V) dC dV$$

$$= U_B(0,0) + U_S(1,1) + \int_0^1 \int_0^1 [\big(1-F(V)\big)g(C) + G(C)f(V)]\, p(V,C) dC dV.$$

(2.7) implies (2.4) and, given (2.3), (2.5). Finally, given (2.5) and that $P_B(\cdot)$and $P_S(\cdot)$ are weakly increasing and decreasing, one can construct a transfer function $x(\cdot, \cdot)$, as in MS (pp. 270-271), such that (p, x) is an incentive-compatible and interim individually rational mechanism. Q.E.D.

MS's Corollary 1. *If traders have positive value densities with overlapping supports, then no incentive-compatible, interim individually rational mechanism can assure ex post efficiency with probability one.*

Proof. Computations that are a special case of those in the proof of Corollary 1 in Section V.C show that the conditions for ex post efficiency with probability one violate (2.5). Q.E.D.

C. *Equilibrium-incentive-efficient trading mechanisms*

Given that ex post efficiency cannot be guaranteed for an incentive-compatible, interim individually rational mechanism, it is natural to consider the limits on efficiency they imply. MS's Theorem 2 addresses this question. Define

(2.8) $$\Phi(V,\alpha) = V - \alpha \frac{1-F(V)}{f(V)} \text{ and } \Gamma(C,\alpha) = C + \alpha \frac{G(C)}{g(C)} \text{ for } \alpha \geq 0,$$

$$p_\alpha(V,C) = 1 \text{ if } \Gamma(C,\alpha) \leq \Phi(V,\alpha), \text{ and } p_\alpha(V,C) = 0 \text{ if } \Gamma(C,\alpha) > \Phi(V,\alpha).$$

MS's Theorem 2. *If there exists an incentive-compatible mechanism (p, x) such that* $U_S(1,1) = U_B(0,0) = 0$ *and* $p = p_\alpha(V,C)$ *for some* $\alpha \in [0,1]$, *then that mechanism maximizes the expected gains from trade among all incentive-compatible, interim individually rational mechanisms. Furthermore, if* $\Phi(V,1)$ *and* $\Gamma(C,1)$ *are increasing on* [0, 1], *then such a mechanism must exist.*

Proof. Note that if feasible, $p_0(V,C)$ would yield an ex post efficient allocation, and $p_1(V,C)$ would maximize the slack in (2.5), which is a kind of "incentive budget constraint". But MS's Corollary 1 shows that $p_0(V,C)$ is unaffordable, and $p_1(V,C)$ wastes surplus. The goal is an optimal compromise between them, balancing the budget while choosing (V, C) combinations on which to trade that yield the largest expected gains per unit of incentive cost. Thus, consider the problem of choosing the function $p(\cdot, \cdot)$ to maximize the expected gains from trade

$$\int_0^1 \int_0^1 (V-C)\, p(V,C) g(C) f(V) dC dV$$

subject to (2.5). If the solution to this problem happens to yield functions $P_B(V)$ and $P_S(C)$ that are monotone increasing and decreasing, respectively, then by MS's Theorem 1, the solution $p(\cdot, \cdot)$ is associated with a mechanism that maximizes the expected gains from trade among all incentive-compatible, interim individually rational mechanisms. Optimality plainly requires $U_S(1,1) = U_B(0,0) = 0$, and that (2.5) holds with equality at the solution. Further, if $\Phi(V, 1)$ and $\Gamma(C, 1)$ are increasing in V and C respectively, then $\Phi(V, \alpha)$ and $\Gamma(C, \alpha)$ are similarly increasing for all $\alpha \in [0,1]$. Thus $p_\alpha(V, C)$, which is defined so that varying α selects the trades that make the greatest contribution to expected gains from trade relative to their unit incentive cost in (2.5), is increasing in V and decreasing in C, and the associated $P_B(V)$ and $P_S(C)$ functions have the required monotonicity properties. Finally, MS (p. 276) show that there always exists an α such that (2.5) holds with equality and $p_\alpha(V, C)$ yields an incentive-compatible mechanism. Q.E.D.

The condition for $p_\alpha(V, C) = 1$, $\Gamma(C, \alpha) \leq \Phi(V, \alpha)$, and $\alpha \geq 0$ imply $V \geq C$, so an equilibrium-incentive-efficient mechanism never requires ex post "perverse" trade. However, as MS (p. 271) note, the transfer function may need to violate ex post individual rationality.

With uniform value densities, MS's Theorem 2 allows a closed-form solution for the incentive-compatible, interim individually rational mechanism that maximizes gains from trade. With uniform densities, (2.8)'s criterion for $p_\alpha(V, C) = 1$, $\Gamma(C, \alpha) \leq \Phi(V, \alpha)$, reduces to

(2.9)
$$v - c \geq \frac{\alpha}{1+\alpha}$$

and (2.4) with equality reduces to

(2.10)
$$0 = \int_{\frac{\alpha}{1+\alpha}}^{1} \int_{0}^{V-\frac{\alpha}{1+\alpha}} (2V - 1 - 2C)\, dC dV = \frac{3\alpha - 1}{6(1+\alpha)^3},$$

which implies that $\alpha = 1/3$ (MS, p. 277).The incentive-efficient direct mechanism then transfers the object if and only if traders' reported values satisfy $v \geq c + 1/4$, at price $(v + c)/3 + 1/6$. With truthful reporting, this outcome function is identical to that of CS's linear double-auction equilibrium: Although the double auction is not incentive-compatible, traders shade their bids in equilibrium to mimic the outcomes of MS's incentive-efficient direct mechanism.[12] The resulting ex ante probability of trade is 9/32 ≈ 28% and the expected total surplus is 9/64 ≈ 0.14.

[12] However, Satterthwaite and Steven Williams (1989, Theorem 5.1) showed that for generic value densities CS's double auction equilibria are incentive-*in*efficient. Thus MS's remarkable result for the case of uniform value densities is a coincidence.

III. A LEVEL-*K* MODEL FOR INCOMPLETE-INFORMATION GAMES

This section specifies a level-*k* model for CS's and MS's trading environment. I focus on the level-*k* model because it is supported by evidence, and for concreteness; but as indicated below many of my results also hold for any nonequilibrium model that makes unique, well-behaved predictions that can be viewed as best responses to some beliefs. I focus on games like those created by direct mechanisms, in which players' decisions are conformable to value estimates, for two reasons. The simplicity of direct mechanisms makes them especially suited to applications. And in more complex, general games, level-*k* models seem less likely to describe people's thinking and evidence to guide a specification for such games is lacking.

Recall that a level-*k* player anchors its beliefs in an *L0* that represents a naive model of other players' responses, with which it assesses the payoff implications of its own decisions before thinking about others' incentives (Crawford et al. 2013, Sections 2.4 and 3). *Lk* then adjusts its beliefs via iterated best responses: *L1* best responds to *L0*, *L2* to *L1*, and so on. In complete-information games, *L0* is normally taken to be uniformly randomly distributed over the range of feasible decisions, and *L1*, *L2*, etc., are defined as iterated best responses. Following Milgrom and Nancy Stokey (1982), Camerer et al. (2004), Crawford and Iriberri (2007), and Crawford et al. (2009), I extend this model to incomplete information by taking *L0*'s decisions as uniform over the feasible decisions and *independent of its own value*.[13] My analysis goes through if the value densities have overlapping supports, provided that *L0* anchors on the overlap. For simplicity, but with loss of generality, I also assume a player's level is independent of its value.

Experiments and some field-data analyses suggest that this generalized level-*k* model gives a unified account of people's non-equilibrium thinking and their informational naiveté, or imperfect attention to how others' decisions depend on their private information. For instance, in an econometric horse race using subjects' initial responses in classic sealed-bid first- and second-price auction experiments, Crawford and Iriberri (2007) showed that, with minor exceptions, this level-*k* model fits better than equilibrium and the leading alternatives. Isabelle Brocas, Juan

[13] Milgrom and Stokey's (1982) notions of Naïve Behavior and First-Order Sophistication, which they suggested might explain zero-sum trades despite their equilibrium no-trade or "Groucho Marx" theorem, are equivalent to an *L1* and *L2* best responding to such an *L0*. It is easy to imagine alternative specifications. An *L0* buyer's bid or seller's ask might be taken to be uniformly distributed below (above) its value, eliminating weakly dominated bids. But *L0* represents not a real player but a player's naïve model of others whose values it doesn't observe. The experiments mentioned next suggest that most people do not perform the contingent reasoning such an *L0* requires. Another alternative model takes *L0* to bid its true expected value, a well-defined notion for direct games. But such a truthful *L0* has much less experimental support in this context (Crawford and Iriberri 2007, Brocas et al. 2014), and it would reduce level-*k*-incentive-compatibility to equilibrium-incentive-compatibility.

Carrillo, Camerer, and Stephanie Wang (2014) reported new experiments in which this level-*k* model explains the patterns in how zero-sum betting deviates from equilibrium. Camerer et al. (2004) showed that a cognitive-hierarchy version of the model explains zero-sum betting in earlier experiments. Finally, Alexander Brown, Camerer, and Dan Lovallo (2012) use this generalized level-*k* model to explain film-goers' failure to draw negative inferences from studios' withholding weak movies from critics before release.

IV. LEVEL-*K* BILATERAL TRADING VIA DOUBLE AUCTION

This section considers bilateral trading via the double auction using the level-*k* model, to set the stage for Section V's analysis of level-*k* design and discussion of level-*k* menu effects. For simplicity I restrict attention to homogeneous populations of *L1*s or *L2*s, which are empirically the most prevalent and illustrate my main points. For *L1*s the analysis applies to general value densities; but for *L2*s I focus on uniform densities. For levels $k = 1, 2$, I denote the buyer's bidding strategy $b_k(V)$ and the seller's asking strategy $a_k(C)$.

A. L1 *traders*

An *L1* buyer believes that the seller's *L0* ask is uniformly distributed on [0, 1]. Thus an *L1* buyer's bid $b_1(V)$ must maximize, over $b \in [0, 1]$

$$\int_0^b \left[V - \frac{a+b}{2}\right] da + \int_b^1 0 \, da.$$

The optimal *L1* strategies are increasing, so the event $a = b$ can again be ignored; and the second-order condition for *L1*'s problem is always satisfied. Solving the first-order condition yields, for any value densities, $b_1(V) = 2V/3$, with range [0, 2/3]. Thus, boundaries aside, an *L1* buyer bids 1/12 more aggressively (bids less) than an equilibrium buyer with uniform value densities: An *L1* buyer's naïve model of the seller systematically underestimates the distribution of the seller's upward-shaded ask, relative to equilibrium, inducing the buyer to underbid.[14]

Similarly, an *L1* seller's ask $a_1(C)$ must maximize, over $a \in [0, 1]$

$$\int_a^1 \frac{a+b}{2} \, db + \int_0^a C \, db.$$

[14] Compare Crawford and Iriberri's (2007) analysis of *L1* and (regarding Section IV.B) *L2* bidding in first-price auctions. An *L1* trader's optimal bidding strategy is independent of the value densities—unlike an *L2*'s, which depends on its partner's density, or an equilibrium trader's, which depends on both densities. Even with multiple equilibria, the level-*k* model makes generically unique predictions, conditional on the empirical parameters that characterize traders' level frequencies.

The first-order condition yields, for any densities, $a_1(C) = 2C/3 + 1/3$, with range [1/3, 1]. An *L1* seller asks 1/12 more aggressively (asks more) than an equilibrium seller with uniform densities.

To sum up, with uniform value densities, *L1* traders' bidding strategies have the same slopes as equilibrium traders' strategies, but are 1/12 more aggressive. When an *L1* buyer meets an *L1* seller, trade takes place if and only if $V \geq C + 1/2$, so the value gap needed for trade is 1/4 larger than for equilibrium traders and ex post efficiency is lost for more value combinations. An *L1* buyer's and seller's ex ante probability of trade is 1/8 = 12.5% and expected surplus is 1/24 ≈ 0.04, far less than the equilibrium probability 9/32 ≈ 28% and expected surplus 9/64 ≈ 0.14 (Section II.B), and even further below the maximum individually rational probability 50% and expected surplus 1/6 ≈ 0.17.

B. L2 *traders*

An *L2* buyer's bid $b_2(V)$ must maximize, over $b \in [0, 1]$

$$\int_0^b \left[V - \frac{a+b}{2}\right] g(a_1^{-1}(a))\, da + \int_b^1 0\, da,$$

where $g(a_1^{-1}(a))$ is the density of an *L1* seller's ask $a_1(C)$ induced by the value density $g(C)$.

If, for instance, $g(C)$ is uniform, an *L2* buyer believes that the seller's ask $a_1(C) = 2C/3 + 1/3$ is uniformly distributed on [1/3, 1], with density 3/2 there and 0 elsewhere. It thus believes that trade requires $b > 1/3$. For $V \leq 1/3$ it is therefore optimal to bid anything it thinks yields 0 probability of trade. In the absence of dominance among such strategies, I set $b_2(V) = V$ for $V \in$ [0, 1/3]. For $V > 1/3$, if $g(C)$ is uniform, an *L2* buyer's bid $b_2(V)$ must maximize over $b \in [1/3, 1]$

$$\int_{1/3}^{\mathrm{b}} \left[V - \frac{a+b}{2}\right] (3/2)\mathrm{d}a.$$

The optimal *L2* strategies are increasing, so the event $a = b$ can again be ignored. The second-order condition is again satisfied. Solving the first-order condition $(3/2)(V - \mathrm{b}) - (3/4)(V - 1/3) = 0$ yields $b_2(V) = 2V/3 + 1/9$ for $V \in [1/3, 1]$, with range [1/3, 7/9].

Comparing an *L2* buyer's optimal strategy to an equilibrium or *L1* buyer's optimal strategy, boundaries aside, with uniform value densities an *L2* buyer bids 1/36 less aggressively (bids more) than an equilibrium buyer, and 1/9 less aggressively than an *L1* buyer: An *L2* buyer's model of the seller systematically overestimates the distribution of the seller's upward-shaded ask, relative to equilibrium, inducing the buyer to overbid.

An *L2* seller's ask $a_2(C)$ must maximize over $a \in [0, 1]$

$$\int_a^1 \frac{a+b}{2} f\big(b_1^{-1}(b)\big)db + \int_0^a C\, f\big(b_1^{-1}(b)\big)db,$$

where $f(b_1^{-1}(b))$ is the density of an *L1* buyer's bid $b_1(V)$ induced by the value density $f(V)$.

If, for instance, $f(V)$ is uniform, an *L2* seller believes that the buyer's bid $b_1(V) = 2V/3$ is uniform on [0, 2/3], with density 3/2 there and 0 elsewhere. It thus believes trade requires $a <$ 2/3. For $C \geq 2/3$ it is therefore optimal for an *L2* seller to bid anything it thinks yields zero probability of trade. In the absence of dominance among such strategies, I set $a_2(C) = C$ for $C \in$ [2/3, 1]. For $C < 2/3$, an *L2* seller's ask $a_2(C)$ must maximize over $a \in [0, 2/3]$

$$\int_a^{\frac{2}{3}} \frac{a+b}{2}(3/2)db + \int_0^a C\,(3/2)db.$$

The second-order condition is satisfied, and the first-order condition $(3/2)(a\text{-}C) + (3/2)(2/3 - C)/2 = 0$ yields $a_2(C) = 2C/3 + 2/9$ for $C \in [0, 2/3]$, with range [2/9, 2/3].

Comparing an *L2* seller's optimal strategy to an equilibrium or *L1* seller's optimal strategy, boundaries aside, with uniform value densities an *L2* seller asks 1/36 less aggressively (asks less) than an equilibrium seller, and 1/9 less aggressively than an *L1* seller.

To sum up, with uniform value densities *L2* traders' strategies again have the same slope as equilibrium traders' strategies, but are 1/36 less aggressive. When an *L2* buyer meets an *L2* seller trade takes place if and only if $V \geq C + 1/6$, so the value gap needed for trade is 1/12 less than for equilibrium traders (1/3 less than for *L1*s), and ex post efficiency is lost for fewer values. The ex ante probability of trade is 25/72 ≈ 35% and expected surplus is 11/72 ≈ 0.15, higher than the equilibrium probability 9/32 ≈ 28% and expected surplus 9/64 ≈ 0.14 (Section II.B), but still well below the maximum individually rational probability 50% and expected surplus 1/6 ≈ 0.17.[15]

V. LEVEL-*K* MECHANISM DESIGN FOR BILATERAL TRADING

This section takes up the design question, replacing equilibrium with a level-*k* model and focusing on direct mechanisms. I focus on the level-*k* model because it is well supported by evidence, and for concreteness; but the proofs show that most of my results also hold for any nonequilibrium model that makes unique predictions that can be viewed as best responses to some beliefs, and which satisfy monotonicity restrictions like those needed for MS's Theorems 1-2. I focus on direct mechanisms because their simplicity is important in applications, and for

[15] When an *L2* buyer meets an *L1* seller, or vice versa, trade occurs when $V \geq C + 1/3$, so the necessary value gap is 1/12 more than for a pair of equilibrium traders and 1/6 more than for a pair of *L2*s, but 1/6 less than for a pair of *L1*s. Because traders' contributions to the value gap are additive, the frequency of trade is determined in general by the population average levels.

them there is clear evidence to guide the specification of a nonequilibrium behavioral model. Finally, as in MS's and almost all design analyses, I ignore decision noise.

I use "incentive-compatible" for direct mechanisms that make it optimal for traders to report truthfully given their beliefs. I derive incentive constraints and the notions of "level-*k*-incentive-compatibility" and "level-*k*-individual-rationality" just as for the usual notions, but for level-*k* beliefs. I call the usual notions "equilibrium-incentive-compatibility" and "equilibrium-interim-individual-rationality". Incentive-efficiency notions are defined for the designer's correct beliefs.

A. *Uniform value densities*

First consider the leading case of uniform value densities, in which MS obtained a closed-form solution for the equilibrium-incentive-efficient direct mechanism, which then mimics CS's linear double-auction equilibrium. Theorem 1 shows that with uniform densities MS's solution is robust to relaxing equilibrium in favor of a level-*k* model, in that MS's equilibrium-incentive-efficient direct mechanism is efficient in the set of level-*k*-incentive-compatible mechanisms (then independent of *k*) for any population of level-*k* traders with $k > 0$, observable or not.

Theorem 1. *With uniform value densities, MS's equilibrium-incentive-efficient direct mechanism is efficient in the set of level-*k*-incentive-compatible mechanisms for any populations of level-*k *traders with k > 0, known or observable or not.*

Proof. With uniform value densities, in the truthful equilibrium of MS's equilibrium-incentive-efficient direct mechanism, each trader faces a uniform distribution of the other's reports. *L1* traders best respond to *L0*s that also imply uniform distributions of reports. The conditions for *L1*-individual rationality and *L1*-incentive-compatibility ((5.2)-5.3) and (5.5)-(5.6) below) then coincide with the analogous conditions for equilibrium traders (my (2.2)-(2.3) and (2.5)-(2.6), MS's (2)-(4)). It follows that MS's equilibrium-incentive-efficient mechanism is also efficient in the set of *L1*-incentive-compatible mechanisms. Further, because MS's mechanism makes *L1*s report truthfully, the conditions for *L2*-individual rationality and *L2*-incentive-compatibility coincide with the conditions for *L1*s, as do *Lk*s' conditions ad infinitum. Thus, MS's equilibrium-incentive-efficient mechanism is also efficient in the set of level-*k*-incentive-compatible mechanisms for any distribution of level-*k* traders with $k > 0$.[16] Q.E.D.

[16] The proof shows that Theorem 1's conclusion holds for any nonequilibrium model in which all traders happen to best respond to *correct* beliefs, as *L1* and higher levels do in the level-*k* model with uniform value densities and uniform *L0*.

B. *Level-k menu effects*

Theorem 1 shows only that with uniform value densities, MS's equilibrium-incentive-efficient mechanism is efficient in the set of *Lk*-incentive-compatible mechanisms if implemented in its *Lk*-incentive-compatible direct form: Section IV's analysis shows that the double auction yields lower surplus for *L1*s than MS's equilibrium-incentive-efficient mechanism, thus violating the revelation principle. Replacing the double auction with MS's mechanism rectifies *L1*s' beliefs, eliminating their aggressiveness and yielding higher surplus. But the double auction preserves *L2*s' beneficial pessimism, yielding higher surplus than in MS's mechanism or any mechanism that is efficient in the set of *L2*-incentive-compatible mechanisms, violating the revelation principle in a different way. The revelation principle fails for level-*k* traders because of Crawford et al.'s (2009) level-*k* menu effects, whereby the choice of mechanism affects the correctness of level-*k* beliefs, which are anchored on an *L0* whose effect is not eliminated by equilibrium thinking. Thus, the choice between mechanisms that are equivalent in an equilibrium analysis is not neutral, and it matters whether level-*k*-incentive-compatibility is required, or can be relaxed to allow traders to respond optimally to any direct mechanism.

C. *General value densities with observable traders' levels*

In this section I continue to assume that level-*k*-incentive-compatibility is required, and consider general well-behaved value densities. For now I further assume that the designer knows or can observe individual traders' levels; and, for ease of notation only, that each trader population is concentrated on one level.[17] While I derive the incentive constraints from traders' level-*k* beliefs, I define the associated incentive-efficiency notion, called "efficiency in the set of level-*k*-incentive-compatible mechanisms", for the designer's correct beliefs.

As in MS's analysis, the only payoff-relevant aspects of a direct mechanism are $p(v, c)$, the probability the object is transferred, and the expected payment $x(v, c)$, where v and c are buyer's and seller's reported values.[18] For any mechanism (p, x), let $f^k(v;p,x)$ and $F^k(v;p,x)$ denote the density and distribution function of an *Lk* seller's beliefs, and $g^k(c;p,x)$ and $G^k(c;p,x)$ the density and distribution function of an *Lk* buyer's. With *L0* uniform random on [0, 1],

[17] Section V.D relaxes the assumption that traders' levels are known or observable, maintaining the restriction to mechanisms that are level-*k*-incentive-compatible for all levels in the trader populations. Section VI relaxes level-*k*-incentive-compatibility to allow any direct mechanism, maintaining the assumption that traders best respond to the chosen mechanism. Throughout the analysis I treat the differences in traders' levels as pure differences of opinion, as in Eliaz and Spiegler (2008): Traders neither believe their partners are better or worse informed, nor draw inferences from their actions or the chosen mechanism.

[18] Thus I assume, following MS, that the mechanism must satisfy ex post (expected) budget balance.

$f^1(v; p, x) \equiv 1$ and $g^1(c; p, x) \equiv 1$. If $\beta_1(V; p, x)$ is an *L1* buyer's response to (*p*, *x*) with value *V* and $\alpha_1(C; p, x)$ is an *L1* seller's response to (*p*, *x*) with cost *C*, $f^2(v; p, x) \equiv f(\beta_1^{-1}(v; p, x))$ and $g^2(c; p, x) \equiv g(\alpha_1^{-1}(c; p, x))$. The beliefs of levels higher than 2 are defined and denoted analogously. I suppress the dependence of $f^2(v; p, x)$ and $g^2(c; p, x)$ on (*p*, *x*) when it is fixed.

Write the buyer's and seller's expected monetary payments, probabilities of trade, and expected gains from trade as functions of their value reports *v* and *c*:

$$X_B^k(v) = \int_0^1 x(v, \hat{c}) g^k(\hat{c}) d\hat{c}, \qquad X_S^k(c) = \int_0^1 x(\hat{v}, c) f^k(\hat{v}) d\hat{v},$$

(5.1) $$P_B^k(v) = \int_0^1 p(v, \hat{c}) g^k(\hat{c}) d\hat{c}, \qquad P_S^k(c) = \int_0^1 p(\hat{v}, c) f^k(\hat{v}) d\hat{v},$$

$$U_B^k(V, v) = V P_B^k(v) - X_B^k(v), \qquad U_S^k(C, c) = X_S^k(c) - C P_S^k(c).$$

For a given *k*, the mechanism *p*(·, ·), *x*(·, ·) is *Lk*-incentive-compatible if and only if truthful reporting is optimal given *Lk* beliefs; that is, for every *V*, *v*, *C*, and *c* in [0, 1],

(5.2) $U_B^k(V, V) \geq U_B^k(V, v) = V P_B^k(v) - X_B^k(v)$ and $U_S^k(C, C) \geq U_S^k(C, c) = X_S^k(c) - C P_S^k(c)$.

Given *Lk*-incentive-compatibility, the mechanism *p*(·, ·), *x*(·, ·) is interim *Lk*-individually rational if and only if, for every *V* and *C* in [0, 1],

(5.3) $$U_B^k(V, V) \geq 0 \text{ and } U_S^k(C, C) \geq 0.$$

Theorems 2 and 3 extend MS's (Theorems 1-2) characterization of equilibrium-incentive-efficient mechanisms to level-*k* models in which the designer knows or can observe traders' levels, so s/he can enforce a different mechanism for each buyer's level *i* and seller's level *j*.

Theorem 2. *Assume that the designer knows or can observe individual traders' levels. Then, for any mechanism that is incentive-compatible for traders of those levels,*

(5.4) $$U_B^i(0,0) + U_S^j(1,1) = \min_{V \in [0,1]} U_B^i(V, V) + \min_{C \in [0,1]} U_S^j(C, C)$$
$$= \int_0^1 \int_0^1 \left(\left[V - \frac{1 - F^j(V)}{f^j(V)} \right] - \left[C + \frac{G^i(C)}{g^i(C)} \right] \right) p(V, C) g^i(C) f^j(V) dC dV.$$

Furthermore, if p(·,·) is any function mapping [0, 1]×[0, 1] *into* [0, 1], *there exists a function x(·,·) such that (p, x) is incentive-compatible and interim individually rational for those levels if and only if* $P_B^i(\cdot)$ *is weakly increasing for all (p, x),* $P_S^j(\cdot)$ *is weakly decreasing for all (p, x), and*

(5.5) $$0 \leq \int_0^1 \int_0^1 \left\{\left[V - \frac{1-F^j(V)}{f^j(V)}\right] - \left[C + \frac{G^i(C)}{g^i(C)}\right]\right\} p(V,C) g^i(C) f^j(V) dC dV.$$[19]

Proof. The proof follows MS's proof, with adjustments for traders' nonequilibrium beliefs. By (5.1), $P_B^i(\cdot)$ is weakly increasing and $P_S^j(\cdot)$ is weakly decreasing for any given (*p*, *x*), which as in MS's proof (pp. 269-270) yields necessary and sufficient conditions for incentive-compatibility:

(5.6) $U_B^i(V,V) = U_B^i(0,0) + \int_0^V P_B^i(\hat{v}) d\hat{v}$ and $U_S^j(C,C) = U_S^j(1,1) + \int_C^1 P_S^j(\hat{c}) d\hat{c}$ for all V and C.

(5.6) implies that $U_B^i(V,V)$ is weakly increasing and $U_S^j(C,C)$ is weakly decreasing, and shows that $U_B^i(0,0) \geq 0$ and $U_S^j(1,1) \geq 0$ suffice for individual rationality for all *V* and *C* as in (5.3).

To derive (5.5), the level-*k* analogue of the equilibrium incentive budget constraint (2.5) (MS's (2)), I follow MS's (p. 270) proof of their Theorem 2, with more detail to make clear that their argument does not depend on the correctness of trader' beliefs. In my notation,

$$\int_0^1 \int_0^1 (V - C)\, p(V,C) g^i(C) f^j(V) dC dV$$

$$= \int_0^1 \int_0^1 [V\, p(V,C) - x(V,C)] g^i(C) dC f^j(V) dV + \int_0^1 \int_0^1 [x(V,C) - C p(V,C)]\, f^j(V) dV g^i(C) dC$$

$$= \int_0^1 [V P_B^i(V) - X_B^i(V)] f^j(V) dV + \int_0^1 [X_S^j(C) - C P_S^j(C)] g^i(C) dC$$

(5.7) $$= \int_0^1 U_B^i(V,V) f^j(V) dV + \int_0^1 U_S^j(C,C) g^i(C) dC$$

$$= U_B^i(0,0) + \int_0^1 \int_0^V P_B^i(\hat{v}) d\hat{v} f^j(V) dV + U_S^j(1,1) + \int_0^1 \int_C^1 P_S^j(\hat{c}) d\hat{c} g^i(C) dC$$

$$= U_B^i(0,0) + U_S^j(1,1) + \int_0^1 [1 - F^j(V)]\, P_B^i(V) dV + \int_0^1 G^i(C)]\, P_S^j(C) dC$$

$$= U_B^i(0,0) + U_S^j(1,1) + \int_0^1 \int_0^1 [G^i(C) f^j(V) + \{1 - F^j(V)\} g^i(C)] p(V,C) dC dV,$$

where the third from last equality follows via integration by parts. Equating the first and last expressions in (5.7) yields (5.4), which implies (5.5) when the mechanism is individually rational. Given (5.3) and that $P_B^j(\cdot)$ is weakly increasing and $P_S^k(\cdot)$ is weakly decreasing, arguments like MS's (pp. 270-271) show that the analogue of their transfer function,

(5.8) $$x(v,c) = \int_0^V v\, d[P_B^i(v)] - \int_0^C c\, d[-P_S^j(c)] + \int_0^1 c[1 - G^i(C)]\, d[-P_S^j(c)],$$

[19] Rene Saran found an important error in my original version of Theorem 2 and showed how to correct it He also noted that the corrected Theorem 2 implies Corollary 1 below, and precludes the ex post "perverse" trade in my previous version. With equilibrium beliefs, $g^i(C;p,x) \equiv g(C)$, $f^j(V;p,x) \equiv f(V)$, and (5.5) is equivalent to MS's (2) incentive budget constraint, (2.5). Because level-*k* beliefs happen to be correct for uniform value densities for all *k*, that equivalence implies Theorem 1.

makes (p, x) incentive-compatible and interim individually rational for traders' levels. Q.E.D.

The analogue of MS's Corollary 1, that no incentive-compatible, individually rational trading mechanism can assure ex post efficiency with probability 1, also goes through for level-k models, with or without the assumption that the designer knows traders' levels.

Corollary 1. *Assume that traders' levels are either homogeneous or heterogeneous, and observable or unobservable to the designer. If the buyer's and seller's values are distributed with positive probability density over [0, 1], then no level-*k*-incentive-compatible and level-*k*-interim individually rational trading mechanism can assure ex post efficiency with probability 1.*

Proof. The proof adapts MS's proof of Corollary 1 (pp. 271-274). For ease of notation, assume first that the designer knows that the buyer is level i and the seller is level j. With positive value densities on [0, 1], when $p(V, C) \equiv 1$ iff $V \geq C$ the incentive budget constraint (5.5) reduces to:

$$0 \leq \int_0^1 \int_0^1 \left\{\left[V - \frac{1 - F^j(V)}{f^j(V)}\right] - \left[C + \frac{G^i(C)}{g^i(C)}\right]\right\} p(V,C) g^i(C) f^j(V) dC dV$$

$$= \int_0^1 \int_0^V [V f^j(V) - 1 + F^j(V)] g^i(C) dC dV - \int_0^1 \int_0^V [C g^i(C) + G^i(C)] dC f^j(V) dV$$

$$= \int_0^1 [V f^j(V) - 1 + F^j(V)] G^j(V) dV - \int_0^1 V G^i(V) f^j(V)] dV \tag{5.9}$$

$$= -\int_0^1 [1 - F^j(V)] G^i(V) dV < 0.$$

This contradiction establishes the result for the case of observable levels. Now assume that the designer knows only the distributions of traders' levels. Level-k-incentive-compatibility then adds cross-level screening constraints to the incentive budget constraints (5.5) conditional on i-j pairs; but adding constraints cannot make the incentive budget constraints feasible. Q.E.D.

Theorem 3, the analogue of MS's Theorem 2, gives a concrete characterization of mechanisms that are efficient in the set of level-k-incentive-compatible mechanisms when the buyer and seller have homogeneous levels, known to the designer. Define, for $\beta \geq 0$,

$$\Psi^{ij}(V, C; \beta) = \left[V - \beta \frac{1 - F^j(V)}{f^j(V)}\right] \left[\frac{g^i(C) f^j(V)}{g(C) f(V)}\right] - \left[C + \beta \frac{G^i(C)}{g^i(C)}\right] \left[\frac{g^i(C) f^j(V)}{g(C) f(V)}\right]$$

$$= (V - C) - \beta \left[\frac{g^i(C) f^j(V)}{g(C) f(V)}\right] \left[\frac{1 - F^j(V)}{f^j(V)} + \frac{G^i(C)}{g^i(C)}\right], \text{ and} \tag{5.10}$$

$$p_\beta^{ij}(V, C) = 1 \text{ if } \Psi^{ij}(V, C; \beta) \geq 0, \text{ and } p_\beta^{ij}(V, C) = 0 \text{ if } \Psi^{ij}(V, C; \beta) \leq 0.$$

If feasible, $p_0^{ij}(V,C)$ would yield an ex post efficient allocation; but Corollary 1 shows that it violates the incentive budget constraint (5.5). By contrast, $p_1^{ij}(V,C)$ maximizes the slack in (5.5); but it wastes surplus. The goal is an optimal compromise between these two extremes.

Theorem 3. *Assume that the designer knows or can observe individual traders' levels,* i *for the buyer and* j *for the seller. If there exists a level*-k-*incentive-compatible mechanism (p, x) such that* $U_B^i(0,0) = U_S^j(1,1) = 0$ and $p = p_\beta^{ij}(V,C)$ *for some* $\beta \in [0,1]$, *then that mechanism maximizes traders' true ex ante expected total gains from trade among all level*-k-*incentive-compatible and level*-k-*interim individually rational mechanisms. Furthermore, if* $\Psi^{ij}(V,C;1)$ *is increasing in V and decreasing in C for any given (p, x), then such a mechanism must exist.*

Proof. The proof adapts MS's proof of their Theorem 2 (pp. 275-276). Fix buyer's and seller's levels *i* and *j*, and consider choosing $p(\cdot,\cdot)$ to maximize traders' ex ante expected total gains from trade subject to $0 \le p(\cdot,\cdot) \le 1$, $U_B^i(0,0) = U_S^j(1,1) = 0$, and (5.5). (5.5) and (5.10) yield:

$$\text{(5.11)} \qquad \max_{\{0 \le p(\cdot,\cdot) \le 1\}} \int_0^1 \int_0^1 (V - C)\, p(V,C) g(C) f(V) dC dV$$

$$\text{s.t. } 0 \le \int_0^1 \int_0^1 \Psi^{ij}(V,C;1) p(V,C) g(C) f(V) dC dV.$$

If a solution $p(\cdot,\cdot)$ to problem (5.11) happens to yield a $\mathrm{P}_\mathrm{B}^\mathrm{i}(\cdot)$ that is weakly increasing for all *v* and a $\mathrm{P}_\mathrm{S}^\mathrm{j}(\cdot)$ that is weakly decreasing for all *c*, then by Theorem 2 that solution is associated with a mechanism that maximizes traders' ex ante expected total gains from trade among all level-*k*-incentive-compatible and level-*k*-interim individually rational mechanisms. (5.11) is like a consumer's budget problem, with the trade probabilities $p(V,C)$ like a continuum of goods and with "prices" $\left\{\left[V - \frac{1-F^j(V)}{f^j(V)}\right] - \left[C + \frac{G^i(C)}{g^i(C)}\right]\right\}$. Because the $p(V,C)$ enter the objective function and the constraint linearly, there are solutions that are "bang-bang", with $p(V,C) = 0$ or 1 almost everywhere and $p(V,C) = 1$ for the (V,C) pairs with the largest expected gain per unit of incentive cost (as for the highest marginal-utility-to-price ratios). By Corollary 1, with *V* and *C* continuously distributed, the constraint holds with equality at a solution. Form the Lagrangean:

$$\int_0^1 \int_0^1 (V - C)\, p(V,C) g(C) f(V) dC dV + \lambda \int_0^1 \int_0^1 \Psi^{ij}(V,C;1) p(V,C) g(C) f(V) dC dV$$

$$\text{(5.12)} \qquad = \int_0^1 \int_0^1 \left(V - C + \lambda \Psi^{ij}(V,C;1)\right) p(V,C) g(C) f(V) dC dV$$

$$= (1+\lambda)\int_0^1\int_0^1 \left(\Psi^{ij}\left(V,C;\frac{\lambda}{1+\lambda}\right)p(V,C)g(C)f(V)dCdV\right).$$

Any function $p(V,C)$ and $\lambda \geq 0$ that satisfy the constraint with equality and the Kuhn-Tucker conditions solves problem (5.11). The Kuhn-Tucker conditions are:

(5.13) $(1+\lambda)\Psi^{ij}\left(V,C;\frac{\lambda}{1+\lambda}\right) \leq 0$ or equivalently $(V-C) - \frac{\lambda}{1+\lambda}\left[\frac{g^i(C)f^j(V)}{g(C)f(V)}\right]\left[\frac{1-F^j(V)}{f^j(V)} + \frac{G^i(C)}{g^i(C)}\right] \leq 0$,

when $p(V,C) = 0$, and

(5.14) $(1+\lambda)\Psi^{ij}\left(V,C;\frac{\lambda}{1+\lambda}\right) \geq 0$ or equivalently $(V-C) - \frac{\lambda}{1+\lambda}\left[\frac{g^i(C)f^j(V)}{g(C)f(V)}\right]\left[\frac{1-F^j(V)}{f^j(V)} + \frac{G^i(C)}{g^i(C)}\right] \geq 0$,

when $p(V,C) = 1$.

An argument as in MS's proof of their Theorem 2, using the continuity and monotonicity of $\Psi^{ij}(V,C;\beta)$, shows that there is a unique λ and $p = p_\beta^{ij}(V,C)$, with $\beta = \frac{\lambda}{1+\lambda}$ (equivalently, $\lambda = \frac{\beta}{1-\beta}$), that satisfy $U_B^i(0,0) = U_S^j(1,1) = 0$, (5.5), (5.13), and (5.14). Q.E.D.

Theorem 3's condition that $\Psi^{ij}(V,C;1)$ is increasing in *V* and decreasing in *C* for all (*p*, *x*) is the level-*k* analogue of MS's Theorem 2 condition that (in my notation from (2.8)) $\Phi(V,1)$ and $\Gamma(C,1)$ are increasing on [0, 1], which is satisfied whenever the true densities fit Myerson's (1981) "regular case", that is ruling out strong hazard rate variations in the "wrong" direction. If traders' beliefs $f^j(V;p,x)$ and $g^i(C;p,x)$ were equal to the true densities $f(V)$ and $g(C)$, Theorem 3's condition would reduce to MS's condition. With level-*k* beliefs, Theorem 3's condition jointly restricts the beliefs and the true densities in a similar but distinct way.

(5.14)'s condition for $p_\beta^{ij}(V,C) = 1$, $\Psi^{ij}(V,C;\beta) \geq 0$, and $\beta \geq 0$ imply that $V \geq C$, so mechanisms that are efficient in the set of level-*k*-incentive-compatible mechanisms, like MS's equilibrium-incentive-efficient mechanisms, never require commitment to ex post perverse trade for any values. However, also as in MS's analysis, the transfer function (5.8) may sometimes violate ex-post individual rationality by requiring payment from buyers who don't get the object.

Theorems 2 and 3 and Corollary 1 show that when the designer can observe traders' levels, so that s/he can enforce a mechanism tailored to each pair of buyer's level *i* and seller's level *j*, MS's characterization of equilibrium-incentive-efficient mechanisms can be fully generalized to a level-*k* model. The proofs show that analogous results go through for any behavioral model that makes unique predictions that are best responses to beliefs and satisfies analogous monotonicity restrictions. Such generalizations are possible even though the latter models can avoid the behaviorally strong assumption that traders play an equilibrium that is a fixed-point in a high-dimensional strategy space, which initiallyappears to play an essential role in MS's analysis.

Comparing the level-*k* incentive budget constraint (5.5) with (2.5) (MS's (2)) and (5.14) with MS's (p. 274) Kuhn-Tucker condition (2.8) shows that the design features that foster equilibrium-incentive-efficiency in MS's analysis also foster efficiency in the set of level-*k*-incentive-compatible mechanisms, but that level-*k* incentives give them different weights.

(5.14) reveals another design feature that fosters efficiency in the set of level-*k*-incentive-compatible mechanisms: Unless such a mechanism happens to induce correct beliefs (as for uniform value densities in Theorem 1), it must benefit from tacit exploitation of predictably incorrect beliefs ("TEPIB"): predictably incorrect in that the level-*k* model predicts traders' deviations from equilibrium; exploitation in the benign sense of using traders' nonequilibrium responses for their own benefit; and tacit in that the mechanism does not actively mislead traders.

A thought-experiment may clarify the influence of TEPIB. Suppose one could exogenously increase the pessimism of traders' level-*k* beliefs, relative to the truth, in the sense of first-order stochastic dominance (moving probability mass higher in the seller's beliefs $f^k(V;p,x)$ and/or lower in the buyer's beliefs $g^k(C;p,x)$). Other things equal, substituting (5.1) into (5.6), that would loosen the incentive budget constraint (5.5) and increase maximized expected total surplus. It is not in fact possible to make traders' beliefs more pessimistic without shifting the true distributions, which also enter the objective function; but pessimism (or optimism) still influences maximized expected surplus, in ways complicated by shifts in the objective function.

Turning to comparative statics, relative to an equilibrium-incentive-efficient mechanism, TEPIB favors trade at (*V*, *C*) combinations whose level-*k* "prices" $\left[\frac{g^i(C)f^j(V)}{g(C)f(V)}\right]\left[\frac{1-F^j(V)}{f^j(V)}+\frac{G^i(C)}{g^i(C)}\right]$ are low compared to their equilibrium prices $\left[\frac{1-F(V)}{f(V)}+\frac{G(C)}{g(C)}\right]$. Finally, for *L2* and higher traders (excluding *L1*s because their beliefs do not depend on the mechanism), TEPIB also favors mechanisms that increase the advantages of the first two effects.

As in MS's analysis, mechanisms that are efficient in the set of level-*k*-incentive-compatible mechanisms can be solved for in closed form only with uniform value densities, for which they happen to induce level-*k* beliefs that are correct, reducing the analysis to an equilibrium one (Theorem 1). To assess the influence of level-*k* thinking, Figure 1 reports the trading regions for mechanisms that are efficient in the set of *L1*-incentive-compatible mechanisms and for equilibrium-incentive-efficient mechanisms, for a comprehensive coarse subset of all possible combinations of linear value densities.[20] Overall, the equilibrium and *L1* trading regions are quite similar. For true densities that make *L1* traders' uniform beliefs pessimistic (optimistic), *L1* trading regions are usually (but not always) supersets (subsets) of equilibrium trading regions.

D. *General value densities with unobservable and heterogeneous traders' levels*

This section relaxes the assumption that each trader population is concentrated on one level, observable by the designer, while continuing to require level-*k*-incentive-compatibility for all levels present with positive probability. Individual traders are assumed to know only their own levels, and the designer knows the support of the level distributions, but possibly nothing more (see footnote 17). In this case Theorem 1 shows that with uniform value densities, even with unknown levels, MS's equilibrium-incentive-efficient direct mechanism is efficient in the set of level-*k*-incentive-compatible mechanisms. Further, Corollary 1 shows that with general value densities that are positive over [0, 1], no level-*k*-incentive-compatible and level-*k*-interim individually rational trading mechanism can assure ex post efficiency with probability 1

Theorem 4 characterizes mechanisms that are efficient in the set of level-*k*-incentive-compatible mechanisms with general value densities and unknown, heterogeneous levels.

A "random posted-price mechanism" is a distribution over prices π and a function $\mu(\cdot)$ such that trade occurs at price π with probability $\mu(\pi)$ if $v \geq \pi \geq c$, with no trade or transfer otherwise.

Theorem 4.[21] *Assume the designer knows that the population level distributions for buyers and sellers include, with positive probabilities,* L1 *and at least one higher level; but s/he cannot observe individual traders' levels. Then, if $f(\cdot)$, $g(\cdot) \neq 1$ almost everywhere, level-*k*-incentive-compatibility requires that the mechanism is equivalent to a random posted-price mechanism. Further, a mechanism that maximizes traders' expected total gains from trade among level-*k*-incentive-compatible and interim individually rational mechanisms is equivalent to a deterministic posted-price mechanism with $U_B^i(0,0) = U_S^j(1,1) = 0$ for all levels i in the buyer*

[20] A few extreme combinations are excluded because they violate the monotonicity conditions that, by Theorems 2 and 3, are needed for the mechanism to be truly optimal. The computations are infeasible for *L2*s, because with $f^2(v) \equiv f(\beta_1^{-1}(v;p,x))$ and $g^2(c) \equiv g(\alpha_1^{-1}(c;p,x))$, (5.5) and (5.14) depend on the transfer function $x(\cdot,\cdot)$ as well as on $p(\cdot,\cdot)$, and the dimensionality of search is too high. The online appendix provides the MATLAB code for *L1*s, written by Rustu Duran.

[21] Rene Saran made an important suggestion and conjecture that led to Theorem 4. A similar result holds, with a similar proof, if the lowest level with positive probability is higher than *L1*; but I give the result as stated for both realism and ease of notation. Theorem 1 shows that if the buyer's and seller's value densities are uniform, the designer has no need to screen traders' levels, and traders' level-*k* incentive constraints for any level are the same as their equilibrium incentive constraints. Theorem 4's assumption that $f(\cdot)$, $g(\cdot) \neq 1$ almost everywhere rules out such cases where *L1*s' beliefs are correct, and trivial variations.

and j in the seller population. The optimal posted price π is unique, characterized by the first-order condition:

(5.15) $$\frac{f(\pi)}{g(\pi)} = \frac{\int_{\pi}^{1}(V-\pi)f(V)dV}{\int_{0}^{\pi}(\pi-C)g(C)dc} = \frac{E(V-\pi|V\geq\pi)}{E(\pi-C|C\leq\pi)}.$$

Proof. Suppose that with positive probabilities, the buyer population includes levels 1 and i, and the seller population includes levels 1 and *j*. By Theorem 2, (5.6) must hold for those levels. If a mechanism is *L1*-incentive-compatible, the proof of Theorem 1 shows that it is *Lk*-incentive-compatible for all *k* > 1 if and only if it is also equilibrium-incentive-compatible. Given (5.1):

(5.16) $U_B^1(V,V) = U_B^1(0{,}0) + \int_0^V P_B^1(\hat{v})d\hat{v} = U_B^1(0{,}0) + \int_0^V \int_0^1 p(\hat{v},\hat{c})d\hat{c}d\hat{v}$ for all *V*,

(5.17) $U_B^i(V,V) = U_B^i(0{,}0) + \int_0^V P_B^i(\hat{v})d\hat{v} = U_B^i(0{,}0) + \int_0^V \int_0^1 p(\hat{v},\hat{c})g(\hat{c})d\hat{c}d\hat{v}$ for all *V*,

(5.18) $U_S^1(C,C) = U_S^1(1{,}1) + \int_C^1 P_S^1(\hat{c})d\hat{c} = U_S^1(1{,}1) + \int_C^1 \int_0^1 p(\hat{v},\hat{c})d\hat{v}d\hat{c}$ for all *C*, and

(5.19) $U_S^i(C,C) = U_S^i(1{,}1) + \int_C^1 P_S^i(\hat{c})d\hat{c} = U_S^i(1{,}1) + \int_C^1 \int_0^1 p(\hat{v},\hat{c})f(\hat{v})d\hat{v}d\hat{c}$ for all *C*.

Standard arguments show that (5.16)-(5.19) are almost everywhere differentiable in *V* or *C*. Because $f(\cdot)$ and g$(\cdot)$ are continuous, differentiability fails only where $p(v,c)$ is discontinuous, hence $p(v,c)$ is almost everywhere continuous. With $f(\cdot)$, g$(\cdot) \neq 1$ almost everywhere, in any open neighborhood throughout which (5.16)-(5.19) are differentiable and $p(v,c)$ is continuous, $p(v,c)$ must be locally independent of *v* and *c*. Given (2.1), (2.6), (5.1), and (5.6), and the utilities buyers of levels 1 and *i* and sellers of levels 1 and *j* derive directly from trading the object, their expected utilities from transfers, given true values *V* and *C*, must be:

(5.20) $U_B^1(0{,}0) + \int_0^V P_B^1(\hat{v})d\hat{v} - VP_B^1(V) = U_B^1(0{,}0) + \int_0^V \int_0^1 p(\hat{v},\hat{c})d\hat{c}\,d\hat{v} - V\int_0^1 p(V,\hat{c})d\hat{c}$,

(5.21) $U_B^i(0{,}0) + \int_0^V P_B^i(\hat{v})d\hat{v} - VP_B^i(V) = U_B^i(0{,}0) + \int_0^V \int_0^1 p(\hat{v},\hat{c})g(\hat{c})d\hat{c}\,d\hat{v} - V\int_0^1 p(V,\hat{c})g(\hat{c})d\hat{c}$,

(5.22) $U_S^1(1{,}1) + \int_C^1 P_S^1(\hat{c})d\hat{c} + CP_S^1(C) = U_S^1(1{,}1) + \int_C^1 \int_0^1 p(\hat{v},\hat{c})d\hat{v}\,d\hat{c} + C\int_0^1 p(\hat{v},C)d\hat{v}$, and

(5.23) $U_S^j(1{,}1) + \int_C^1 P_S^j(\hat{c})d\hat{c} + CP_S^j(C) = U_S^j(1{,}1) + \int_C^1 \int_0^1 p(\hat{v},\hat{c})f^j(\hat{v})d\hat{v}\,d\hat{c} + C\int_0^1 p(\hat{v},C)f^j(\hat{v})d\hat{v}$.

For (5.20)-(5.23) to hold for all *V* and *C*, the derivatives of (5.20) and (5.21) with respect to *V* and the derivatives of (5.22) and (5.23) with respect to *C* must each be equal whenever they exist. Thus, after cancellations we must have (with subscripts denoting partial differentiation):

(5.24) $\int_0^1 p_1(\hat{v},\hat{c})d\hat{c} = \int_0^1 p_1(\hat{v},\hat{c})g(\hat{c})d\hat{c}$ and $\int_0^1 p_2(\hat{v},\hat{c})d\hat{c} = \int_0^1 p_2(\hat{v},\hat{c})g(\hat{c})d\hat{c}$.

It follows that with $f(\cdot)$, g$(\cdot) \neq 1$ almost everywhere, $p(v,c)$ must be equivalent to a random posted-price mechanism. Any random posted-price mechanism is level-*k*-incentive-compatible for any k. Because the $p(v,c)$ enter the objective function and constraints linearly, there are always solutions that are "bang-bang", with $p(V,C) = 0$ or 1 almost everywhere, so $\mu(\cdot) \equiv 1$ without loss of generality. Finally, an optimal deterministic posted-price mechanism π solves:

$$(5.25) \qquad \max_{\{0 \le \pi \le 1\}} \int_{\pi}^{1} \int_{0}^{\pi} (V - C)\, p(V,C) g(C) f(V) dC dV.$$

The second-order condition of problem (5.25) is satisfied globally, so there is a uniquely optimal posted price, which is characterized by the first-order condition (5.15). Q.E.D.

Theorem 4 shows that the designer's need to screen traders' levels of thinking as well as their values rules out the sensitive dependence on reported values of mechanisms that are equilibrium-incentive-efficient or efficient in the set of level-*k*-incentive-compatible mechanisms when traders' levels are known or observable. This result gives a plausible, more basic rationale for the distribution-free, dominant-strategy implementation often assumed in the literature on robust mechanism design (Hagerty and Rogerson 1987; Čopič and Ponsatí 2008, 2015).

The optimal posted price characterized by (5.15) balances the expected gains and losses in the values for which trade occurs from changing π. The optimal price is determined by the true expected-surplus tradeoffs, independent of traders' nonequilibrium beliefs, so there are no TEPIB benefits. The case of approximately uniform value densities gives a rough idea of how much surplus is lost by giving up sensitive dependence on reported values (even though with exactly uniform densities the designer is not restricted to a posted-price mechanism by Theorem 1): MS's equilibrium-incentive-efficient mechanism then yields approximately an ex ante probability of trade of 9/32 ≈ 28% and expected surplus of 9/64 ≈ 0.14. The optimal posted price with uniform densities is ½, which would yield ex ante probability of trade of 1/4 = 25% and expected surplus of 1/8 = 0.125, a modest cost for the added robustness of a posted price.

Theorem 4's result continues to hold for any sufficiently well-behaved nonequilibrium model where strategic thinking falls into identifiable classes, and which given a class, makes unique predictions that can be viewed as best responses to some beliefs. The resulting static posted-price mechanisms come closer to satisfying Wilson's (1987) desideratum, in that their rules are distribution-free. However, the optimal posted price is determined, via (5.15), by conditional means that depend on the full value densities. Čopič and Ponsatí (2015) note that a dynamic implementation can avoid such dependence, in that any (even random) posted-price mechanism can be implemented via a continuous-time double auction in which the auctioneer reveals bids to traders only once they are compatible. This fully satisfies Wilson's desideratum, in that it does not require the designer to know the value densities and comes closer to real-world mechanisms.

VI. RELAXING LEVEL-*K*-INCENTIVE-COMPATIBILITY

This section relaxes Section V's requirement of level-*k*-incentive-compatibility, instead allowing any direct mechanism. Here one can still define a general class of feasible direct mechanisms, with payoff-relevant outcomes $p(v, c)$ and $x(v, c)$. However, a mechanism's incentive effects can no longer be tractably captured via constraints like (5.2) and (5.6), but must be modeled directly via level-*k* traders' responses to it. When level-*k*-incentive-compatibility is not required, I call a mechanism "level-*k*-incentive-efficient" if its outcomes cannot be improved upon by any direct mechanism. I first consider the case where traders' levels are known or observable by the designer, and then the case of unknown or unobservable and heterogeneous traders' levels.

A. *Observable traders' levels*

With observable traders' levels, assume uniform value densities for simplicity, and as a tractable proxy for what is achievable via any direct mechanism, consider double auctions with reserve prices chosen by the designer. Reserve prices have no benefits if *Lk* traders continue to anchor beliefs on an *L0* uniform random on the full range of possible values [0, 1]. But a double auction with a restricted menu of bids or asks may make level-*k* traders anchor on the restricted menu instead of [0, 1], and such anchoring can make reserve prices useful.[22]

For example, *L1* traders in a double auction without reserve prices believe they face bids or asks uniformly distributed on [0, 1], yielding *L1*-incentive-*in*efficient outcomes. In a double auction with reserve prices for buyer's bids of ¾ and seller's asks of ¼, if *L1* traders anchor on the restricted menu, they will bid or ask as if they faced asks or bids uniformly distributed on [¼, 1] or [0, ¾] respectively, or equivalently (given the ranges of their optimal bids or asks) on [¼, ¾] for both: exactly the ranges of serious bids or asks in CS's linear double-auction equilibrium (Section IV.A). A double auction with those reserve prices therefore rectifies *L1* traders' beliefs, and yields the same outcomes as MS's equilibrium-incentive-efficient direct mechanism: ex ante probability of trade $9/32 \approx 28\%$ and expected surplus $9/64 \approx 0.14$ (Section II.B), far higher than *L1*s' ex ante probability of trade $1/8 = 12.5\%$ and expected surplus $1/24 \approx 0.04$ in the double

[22] I know of no evidence for (or against) such an *L0* specification, but in marketing analogous menu effects are commonplace MS's general specification of feasible mechanisms implicitly allows reserve prices, and their analysis therefore shows that if equilibrium is assumed, reserve prices are not useful in this setting. Crawford et al. (2009) showed that in first-price auctions, level-*k* bidders anchoring on a restricted menu of bids can make reserve prices useful even when they would be useless with equilibrium bidders. Saran (2011b) studies how menu-dependent preferences affect the revelation principle.

auction without reserve prices. Pushing the reserve prices further than ¾ and ¼ further reduces the value gap needed for trade, other things equal a benefit; but it also begins to preclude some bids or asks needed for trade. Computations show that the cost of precluding such bids or asks exceeds the benefits, and that reserve prices of ¾ and ¼ are in fact optimal for the designer.

For *L2*s, a double auction without reserve prices already improves upon MS's equilibrium-incentive-efficient mechanism or a mechanism that is efficient in the set of *L2*-incentive-efficient mechanisms (Sections II.B, IV.B, and V.A). Feasible reserve prices (that is, restricted to [0, 1]) bring *L2* traders' beliefs closer to their equilibrium beliefs, eliminating some of the beneficial pessimism that allows the double auction without reserve prices to yield better outcomes for them. Computations show that a double auction without reserve prices is in fact optimal. It has ex ante probability of trade 25/72 ≈ 35% and expected surplus 11/72 ≈ 0.15, higher than the equilibrium probability of trade 9/32 ≈ 28% and expected surplus 9/64 ≈ 0.14 (Section II.B).

B. *Unobservable and heterogeneous traders' levels*

Turning to the case of unobservable and heterogeneous traders' levels, Section V.C-D's results allow a simple rough estimate of the potential benefits of allowing non-level-*k*-incentive-compatible direct mechanisms. Suppose, for example, the designer knows that the population includes multiple levels with positive probability but that the frequency of one level is very high. Then a mechanism that is efficient in the set of level-*k*-incentive-compatible mechanisms for only the level with high frequency, or possibly a non-level-*k*-incentive-compatible mechanism as discussed in Section VI.A, will generally improve upon a mechanism that is efficient in the set of level-*k*-incentive-compatible mechanisms for the actual population of levels.

For example, with uniform value densities the mechanism that is efficient in the set of level-*k*-incentive-compatible mechanisms is a posted-price mechanism, with optimal posted price ½, ex ante probability of trade 1/4 = 25%, and expected total surplus 1/8 = 0.125. By contrast, a mechanism that is efficient in the set of *L1*-incentive-compatible mechanisms with uniform value densities yields ex ante probability of trade 9/32 ≈ 28% and expected surplus 9/64 ≈ 0.14 (Section II.B). Thus, even though the former estimate is only an approximation for approximately uniform densities, when the frequency of *L1* is high switching from the former mechanism to the latter must yield a significant gain, even accounting for other levels that are allowed to violate level-*k*-incentive-compatibility. And when the frequency of *L2* is high, the *L2*-

incentive-efficient double auction without reserve prices yields ex ante probability of trade 25/72 ≈ 35% and expected surplus 11/72 ≈ 0.15, an even more significant approximate gain.

More generally, relaxing the restriction to level-*k*-incentive-compatible mechanisms can yield level-*k*-incentive-efficient mechanisms that differ qualitatively as well as quantitatively from equilibrium-incentive-efficient mechanisms, with substantial gains in incentive-efficiency.

VII. CONCLUSION

A structural nonequilibrium level-*k* model makes specific predictions that allow an analysis of mechanism design for bilateral trading with power comparable to MS's classic equilibrium analysis. The results clarify the role of the equilibrium assumption in several ways.

The anchoring of level-*k* beliefs on *L0* creates level-*k* menu effects that make the revelation principle fail. It thus matters whether level-*k*-incentive-compatibility is truly required.

If level-*k*-incentive-compatibility is required, MS's analysis is surprisingly robust to relaxing equilibrium in favor of a level-*k* or another kind of structural nonequilibrium model. MS's closed-form solution for the equilibrium-incentive-efficient mechanism with uniform value densities remains valid for any level-*k* model—though only if that mechanism is implemented in its level-*k*-incentive-compatible form, not as the double auction.

Further, for general well-behaved value densities with traders' levels known or observable by the designer, MS's characterization of incentive-efficient mechanisms is fully robust to level-*k* thinking, with one novel feature, tacit exploitation of predictably incorrect beliefs ("TEPIB"). In that case mechanisms that are efficient in the set of level-*k*-incentive-compatible mechanisms are just as sensitive to the details of the environment as equilibrium-incentive-efficient mechanisms.

By contrast, for general well-behaved value densities with traders' levels heterogeneous and unobservable by the designer, who must then screen traders' levels along with their values, level-*k*-incentive-compatibility compels the use of a random posted-price mechanism, and mechanisms that are efficient in the set of level-*k*-incentive-compatible mechanisms are deterministic posted-price mechanisms. Such mechanisms neutralize TEPIB, but their reduced sensitivity to environmental details brings them closer to real institutions, as Wilson (1987) advocated.

In either case, if level-*k*-incentive-compatibility is required, MS's Corollary 1 result that no incentive-compatible, interim individually rational mechanism can assure ex-post efficiency is fully robust to level-*k* thinking.

The proofs of my results for the case where level-*k*-incentive-compatibility is required show that an analysis like MS's can dispense with the assumption that traders play an equilibrium, or even the special structure of level-*k* models: Those results hold as long as traders follow well-behaved decision rules that yield unique predictions that can be viewed as best responses to some beliefs, and which are known or observable by the designer. Thus in this case the force of the equilibrium assumption comes not from equilibrium per se, but from its implications that traders best respond to some beliefs, and that the designer knows individual traders' decision rules.

The power of the equilibrium assumption also stems from its implication, via the revelation principle, that it does not matter whether incentive-compatibility is truly required. If *non*-level-*k*-incentive-compatible direct mechanisms are truly feasible, it can be beneficial to violate incentive-compatibility in various ways that are irrelevant in an equilibrium analysis.

Suppose, for instance, that traders' levels are known or observable by the designer, traders' value densities are uniform, and menus of bids or asks restricted by reserve prices make level-*k* traders anchor on the restricted menu instead of on [0, 1]. Then, for *L1*s, reserve prices chosen optimally by the designer allow the double auction to rectify their beliefs and mimic the outcome of MS's equilibrium-incentive-efficient mechanism. For *L2*s, by contrast, the double auction *without* reserve prices is optimal, and it improves upon any mechanism that is efficient in the set of *L2*-incentive-efficient mechanisms, or MS's equilibrium-incentive-efficient mechanism.

If, instead, traders' levels are heterogeneous and unobservable by the designer, but the population frequency of one level is known to be high, the designer can benefit by violating incentive-compatibility for other levels in favor of a level-*k*-incentive-compatible mechanism for the level whose frequency is high, or possibly a non-level-*k*-incentive-compatible mechanism.

More generally, relaxing level-*k*-incentive-compatibility, if feasible, can yield level-*k*-incentive-efficient mechanisms that differ qualitatively as well as quantitatively from equilibrium-incentive-efficient mechanisms, sometimes with large efficiency gains.

It is my hope that this paper's analysis shows that a nonequilibrium analysis of mechanism design can add new insights, and that it will encourage further progress in that direction.

REFERENCES

Abdulkadiroglu, Atila, Yeon-Koo Che, and Yosuke Yasuda (2011): "Resolving Conflicting Preferences in School Choice: The 'Boston Mechanism' Reconsidered," *American Economic Review*, 101, 1–14.

Abdulkadiroglu, Atila, and Tayfun Sönmez (2003): "School Choice: A Mechanism Design Approach," *American Economic Review*, 93, 729-747.

Aumann, Robert J., and Adam Brandenburger (1995): "Epistemic Conditions for Nash Equilibrium," *Econometrica*, 63, 1161-1180.

Bergemann, Dirk, Stephen Morris, and Olivier Tercieux (2011): "Rationalizable Implementation," *Journal of Economic Theory*, 146, 1253-1274.

Bernheim, B. Douglas (1984): "Rationalizable Strategic Behavior," *Econometrica*, 52, 1007-1028.

Brocas, Isabelle, Juan D. Carrillo, Colin F. Camerer, and Stephanie W. Wang (2014): "Imperfect Choice or Imperfect Attention? Understanding Strategic Thinking in Private Information Games," *Review of Economic Studies*, 81, 944-970.

Brown, Alexander, Colin F. Camerer, and Dan Lovallo (2012): "To Review or Not To Review? Limited Strategic Thinking at the Movie Box Office." *American Economic Journal: Microeconomics*, 4, 1-28.

Bulow, Jeremy, and John Roberts (1989): "The Simple Economics of Optimal Auctions," *Journal of Political Economy*, 97, 1060-1090.

Camerer, Colin F., Teck-Hua Ho, and Juin Kuan Chong (2004): "A Cognitive Hierarchy Model of Games," *Quarterly Journal of Economics*, 119, 861-898.

Chatterjee, Kalyan, and William Samuelson (1983): "Bargaining under Incomplete Information," *Operations Research*, 31, 835-851.

Chen, Yan, and John O. Ledyard (2008): "Mechanism Design Experiments," in *The New Palgrave Dictionary of Economics, Second Edition*, ed. Steven Durlauf and Lawrence Blume. London: Macmillan.

Čopič, Jernej, and Clara Ponsatí (2008): "Robust Bilateral Trade and Mediated Bargaining," *Journal of the European Economic Association*, 6, 570-589.

Čopič, Jernej, and Clara Ponsati (2015): "Optimal Robust Bilateral Trade: Risk Neutrality"; http://chopichcode.org/assets/Copic_Ponsati_2015-Risk-neutral.pdf .

Crawford, Vincent P., Miguel A. Costa-Gomes, and Nagore Iriberri (2013): "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications," *Journal of Economic Literature*, 51, 5-62.

Crawford, Vincent P., and Nagore Iriberri (2007): "Level-*k* Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?," *Econometrica*, 75, 1721-1770.

Crawford, Vincent P., Tamar Kugler, Zvika Neeman, and Ady Pauzner (2009): "Behaviorally Optimal Auction Design: An Example and Some Observations," *Journal of the European Economic Association*, 7, 365-376.

de Clippel, Geoffroy, Rene Saran, and Roberto Serrano (2015): "Mechanism Design with Bounded Depth of Reasoning and Small Modeling Mistakes"; https://drive.google.com/file/d/0B8Sv4TBdx30Jc0JRRUpkb3EzVTQ/edit?usp=sharing .

Eliaz, Kfir, and Ran Spiegler (2006): "Contracting with Diversely Naïve Agents," *Review of Economic Studies*, 73, 689–714.

Eliaz, Kfir, and Ran Spiegler (2007): "A Mechanism-Design Approach to Speculative Trade," *Econometrica*, 75, 875–884.

Eliaz, Kfir, and Ran Spiegler (2008): "Consumer Optimism and Price Discrimination," *Theoretical Economics*, 3, 459–497.

Erdil, Aytek, and Haluk Ergin (2008): "What's the Matter with Tie-Breaking? Improving Efficiency in School Choice," *American Economic Review*, 98, 669-689.

Glazer, Jacob, and Ariel Rubinstein (1998): "Motives and Implementation: On the Design of Mechanisms to Elicit Opinions," *Journal of Economic Theory*, 79, 157–173.

Gorelkina, Olga (2015): "The Expected Externality Mechanism in a Level-*k* Environment," working paper, Max Planck Institute, Bonn; http://ssrn.com/abstract=2550085

Hagerty, Kathleen M., and William P. Rogerson (1987): "Robust Trading Mechanisms," *Journal of Economic Theory*, 42, 94-107.

Healy, Paul J. (2006): "Learning Dynamics for Mechanism Design: An Experimental Comparison of Public Goods Mechanisms," *Journal of Economic Theory*, 129, 114 – 149.

Katok, Elena, Martin Sefton, and Abdullah Yavas (2002): "Implementation by Iterative Dominance and Backward Induction: An Experimental Comparison," *Journal of Economic Theory*, 104, 89–103.

Kneeland, Terri (2013): “Mechanism Design with Level-*k* Types,” manuscript.

Maskin, Eric (2011): “Commentary: Nash Equilibrium and Mechanism Design,” *Games and Economic Behavior*, 71, 9-11.

Matsushima, Hitoshi (2007): “Mechanism Design with Side Payments: Individual Rationality and Iterative Dominance,” *Journal of Economic Theory*, 133, 1– 30.

Matsushima, Hitoshi (2008): “Detail-free Mechanism Design in Twice Iterative Dominance: Large Economies,” *Journal of Economic Theory*, 141, 134–151.

Milgrom, Paul, Lawrence Ausubel, Jon Levin, and Ilya Segal (2012). “Incentive Auction Rules Option and Discussion,” report to the Federal Communications Commission by Auctionomics and Power Auctions; https://apps.fcc.gov/edocs_public/attachmatch/FCC-12-118A2.pdf

Milgrom, Paul R., and Nancy Stokey (1982): “Information, Trade and Common Knowledge,” *Journal of Economic Theory*, 26, 17-26.

Milgrom, Paul, and Robert Weber (1982): “A Theory of Auctions and Competitive Bidding,” *Econometrica*, 50, 1089-1122.

Mookherjee, Dilip, and Stefan Reichelstein (1992): “Dominant Strategy Implementation of Bayesian Incentive Compatible Allocation Rules,” *Journal of Economic Theory*, 56, 378-399.

Myerson, Roger B. (1981): “Optimal Auction Design,” *Mathematics of Operations Research*, 6, 58-73.

Myerson, Roger B., and Mark A. Satterthwaite (1983): “Efficient Mechanisms for Bilateral Trading,” *Journal of Economic Theory*, 29, 265–81.

Neeman, Zvika (2003): “The Effectiveness of English Auctions,” *Games and Economic Behavior*, 43, 214–238.

Saran, Rene (2011a): “Bilateral Trading with Naive Traders,” *Games and Economic Behavior*, 72, 544–557.

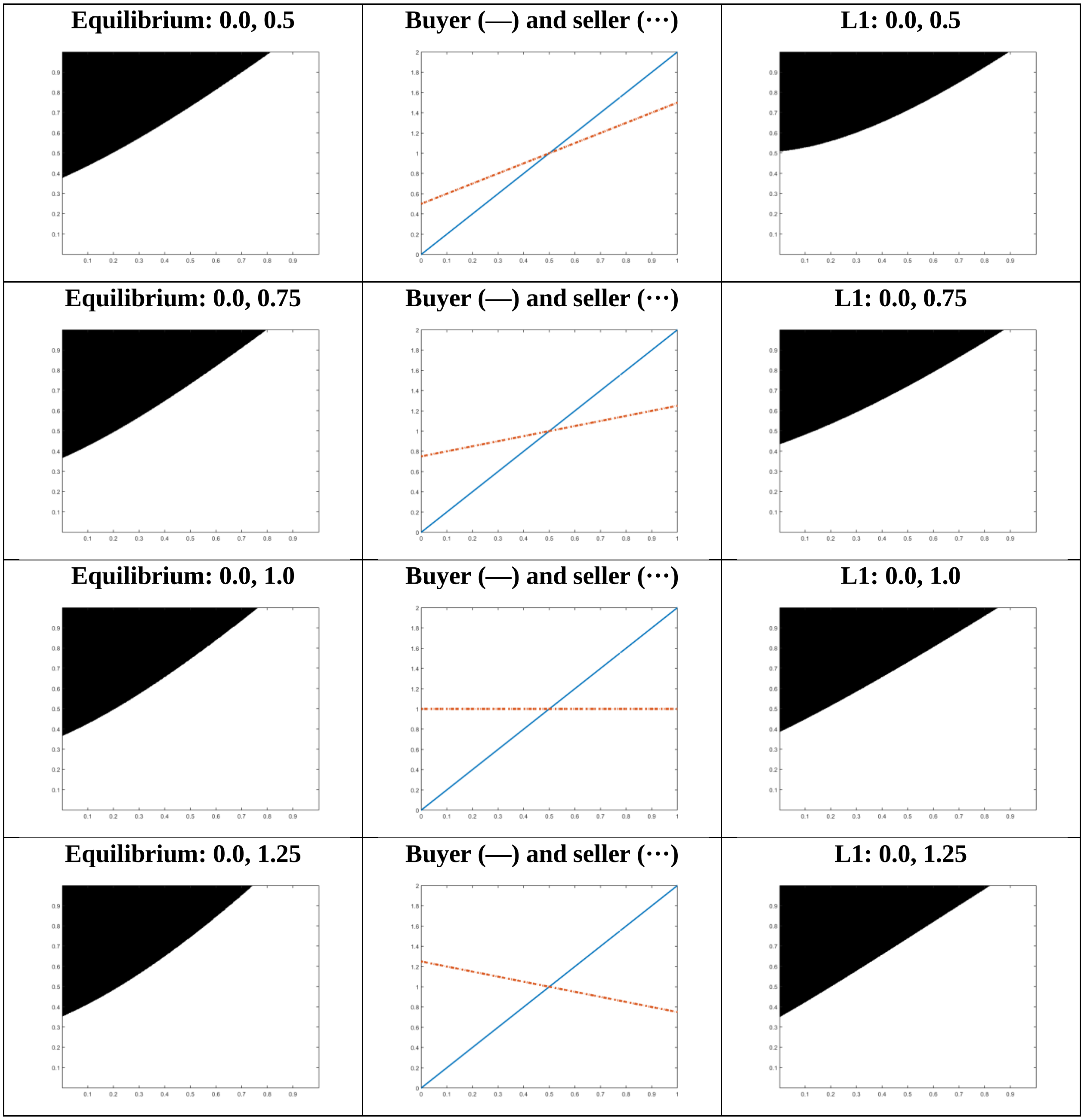
Saran, Rene (2011b): “Menu-dependent Preferences and Revelation Principle,” *Journal of Economic Theory*, 146, 1712-1720.

Satterthwaite, Mark, and Steven R. Williams (1989): “Bilateral Trade with the Sealed Bid k-Double Auction: Existence and Efficiency,” *Journal of Economic Theory*, 48, 107–133.

Wilson, Robert B. (1987): "Game-Theoretic Analyses of Trading Processes," in: *Advances in Economic Theory: Fifth World Congress*, Truman Bewley (ed.); Chapter 2, 33-70, Cambridge UK: Cambridge University Press.

Wolitzky, Alexander (2014): "Mechanism Design with Maxmin Agents: Theory and an Application to Bilateral Trade," manuscript, Stanford University; http://www.stanford.edu/~wolitzky/research/working_papers_pdf/maxmin%20April%202014%20v1.pdf

Figure 1. Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of *L1*-incentive-compatible mechanisms[23]



[23] Buyer's value V is on the vertical axis; seller's value C is on the horizontal axis. All value densities are linear; "x, y" means the buyer's density $f(V)$ satisfies $f(0) = x$ and $f(1) = 2\text{-}x$, and the seller's density $g(C)$ satisfies $g(0) = y$ and $g(1) = 2\text{-}y$.

Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of *L1*-incentive-compatible mechanisms

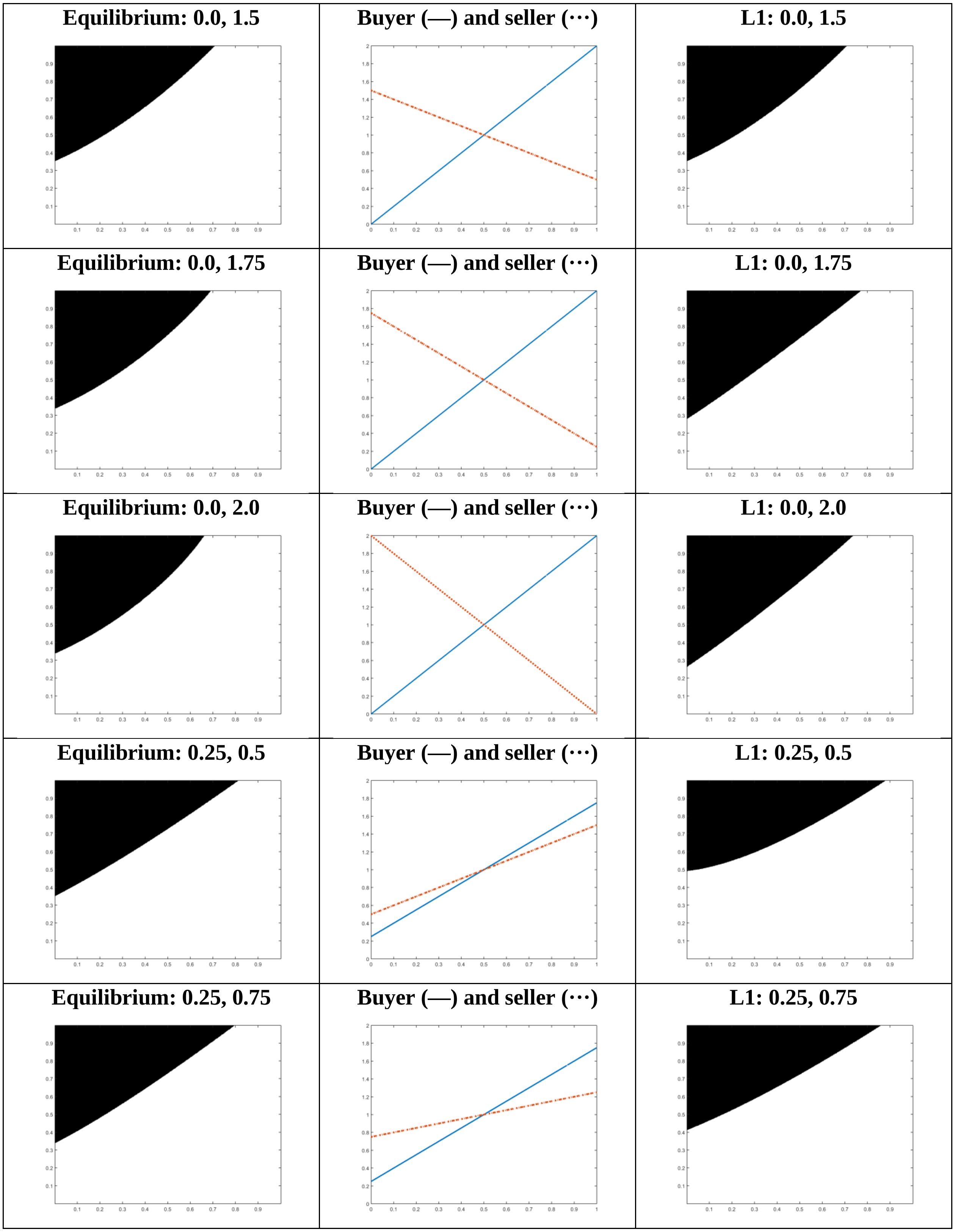


Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of *L1*-incentive-compatible mechanisms

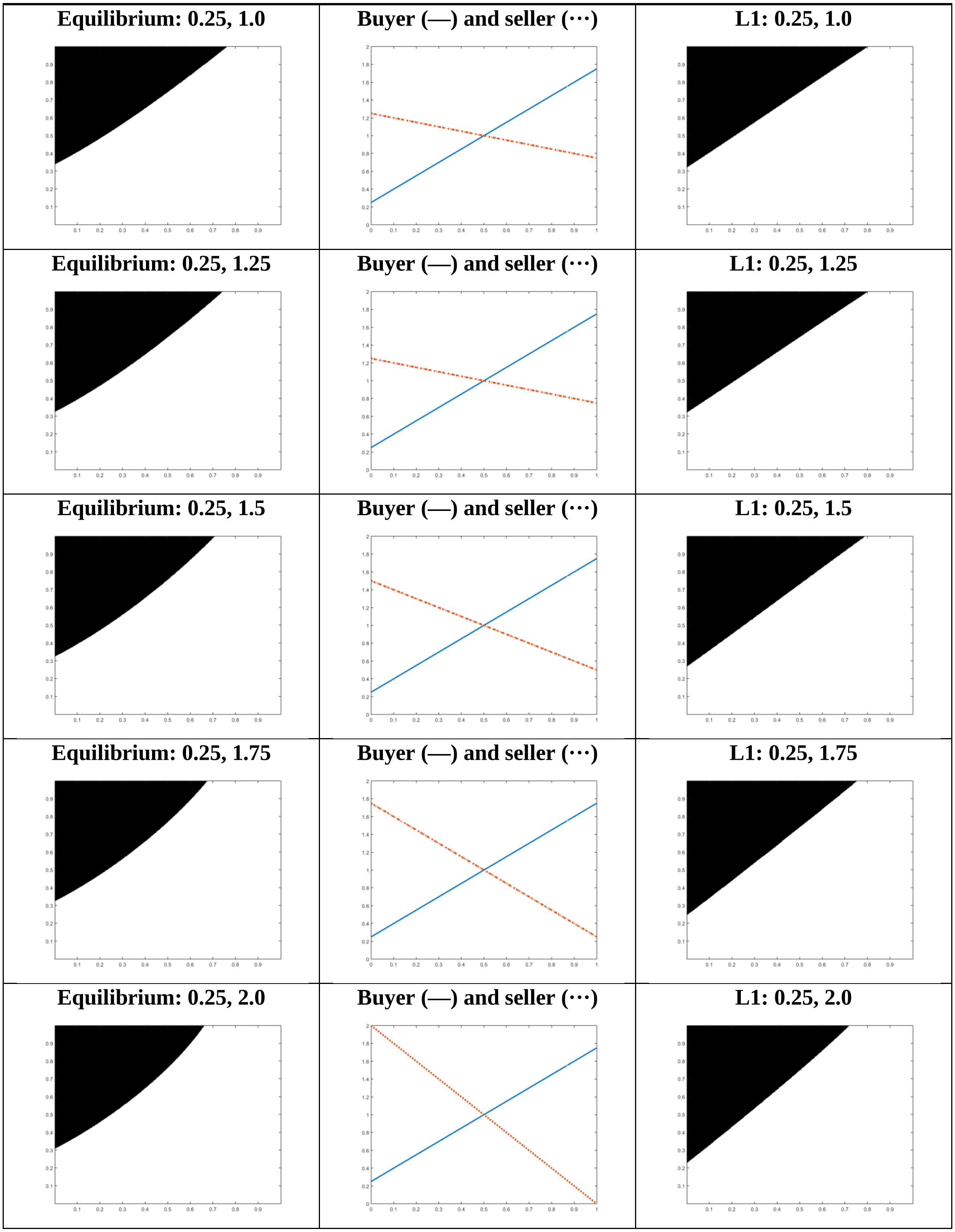


Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of *L1*-incentive-compatible mechanisms

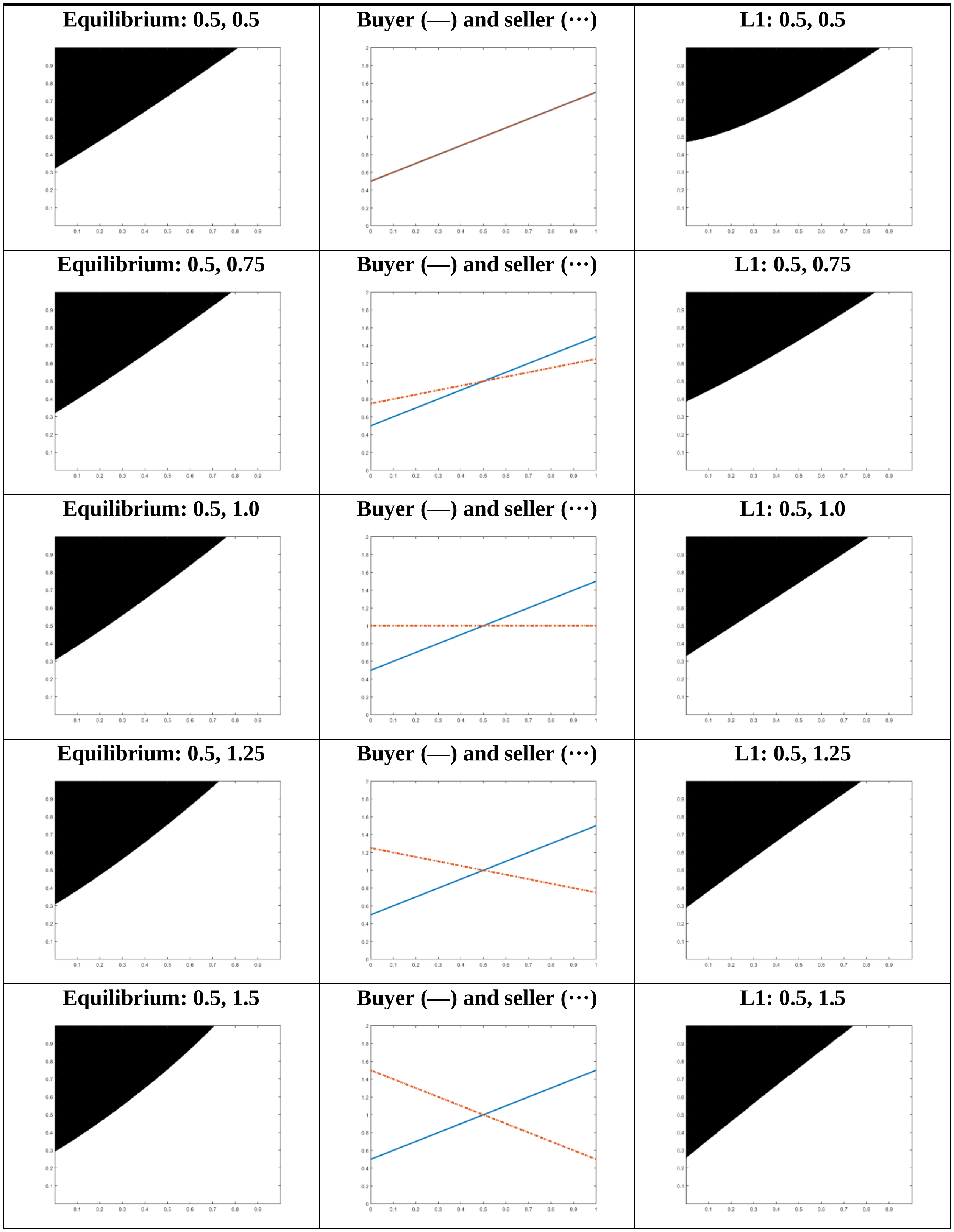


Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of *L1*-incentive-compatible mechanisms

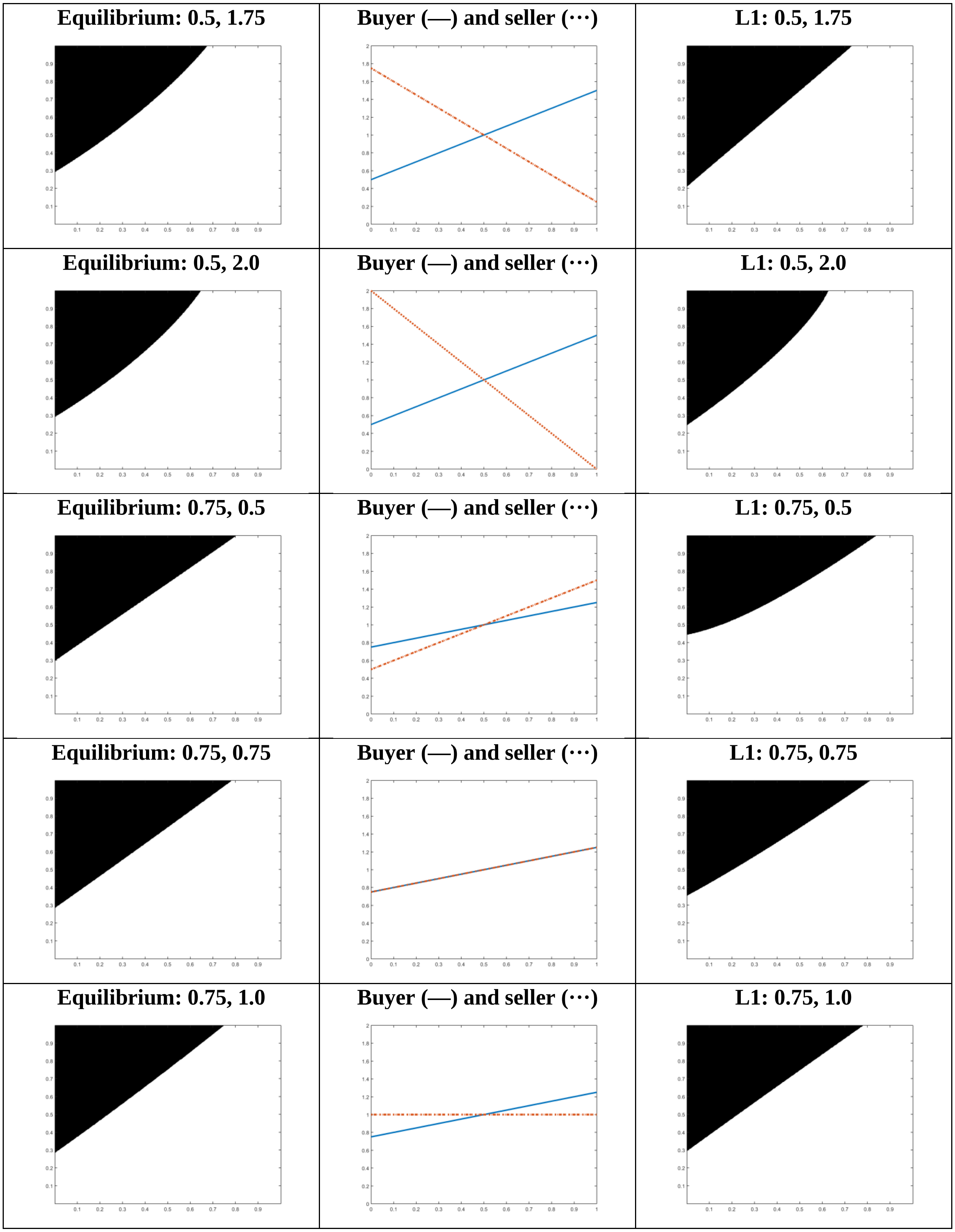


Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of *L1*-incentive-compatible mechanisms

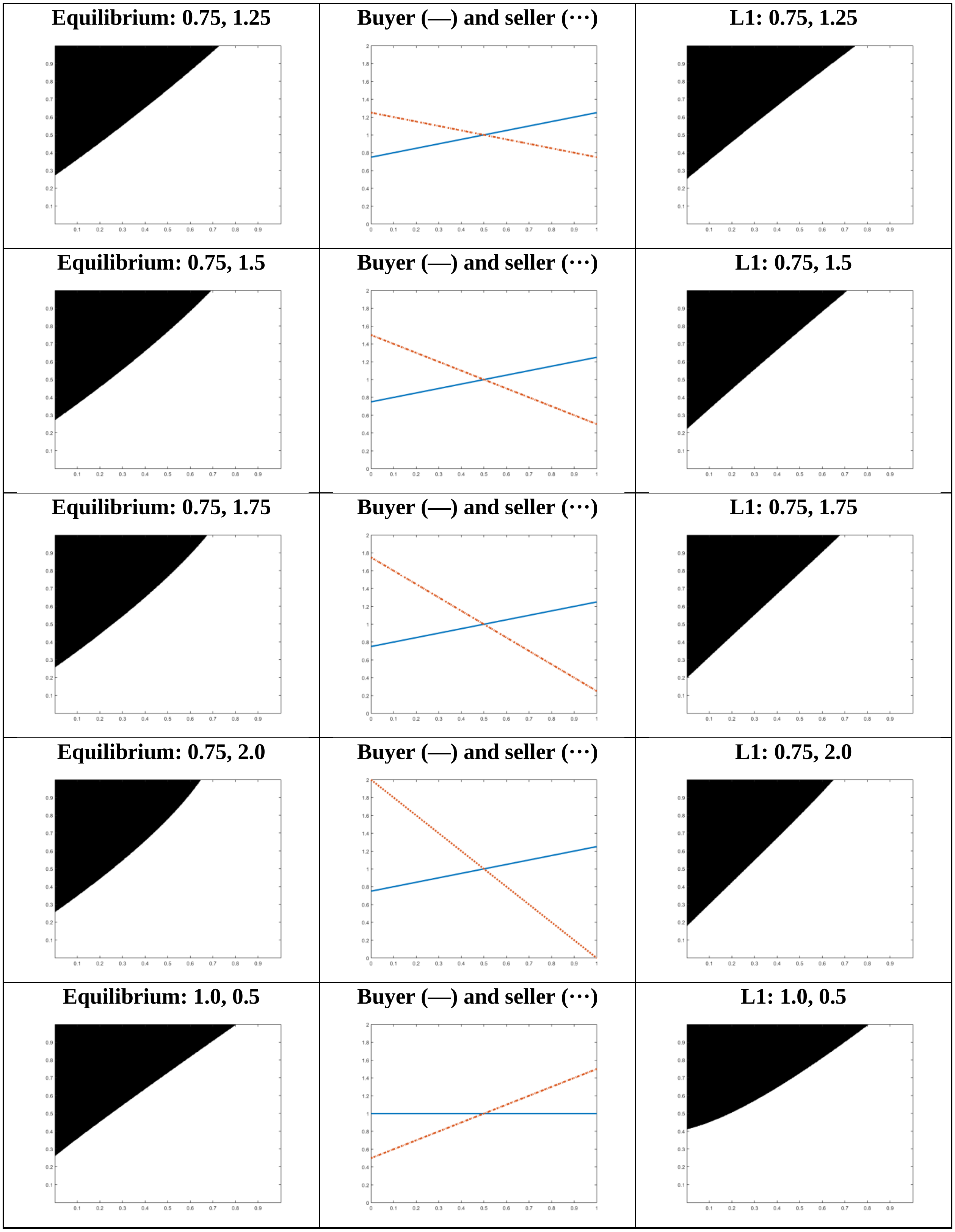


Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of *L1*-incentive-compatible mechanisms

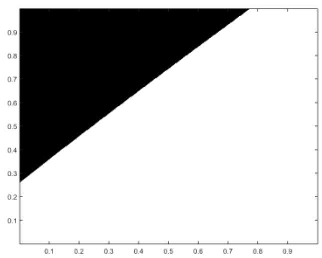
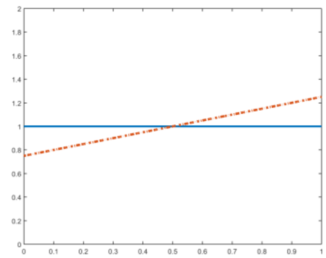
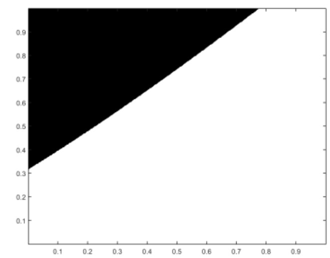
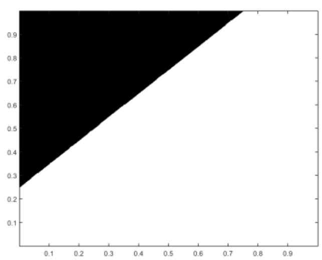
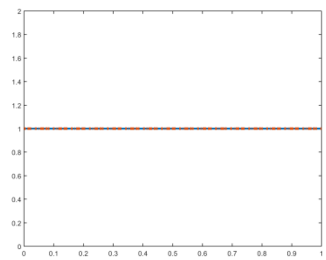
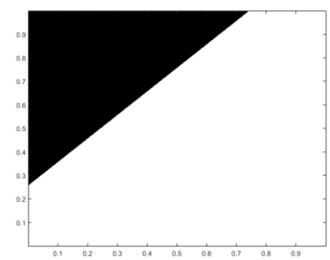
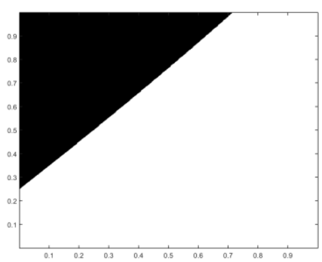
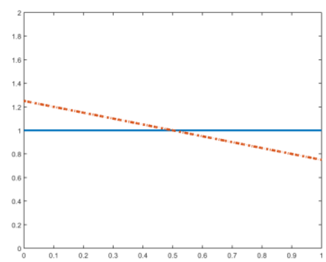
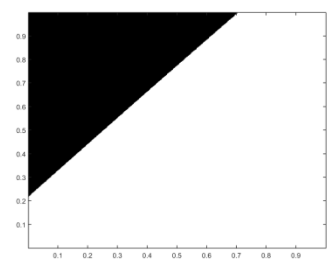
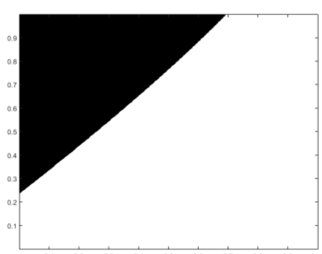
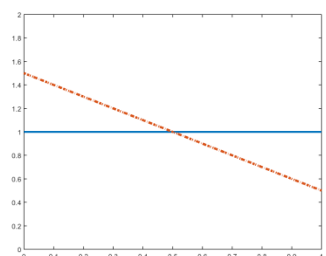
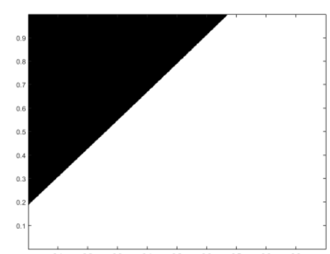
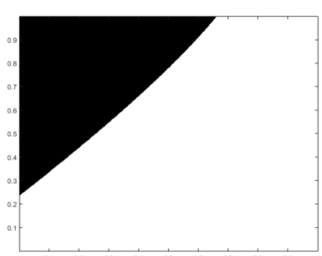
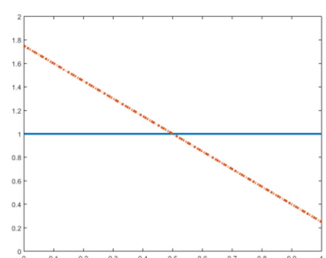
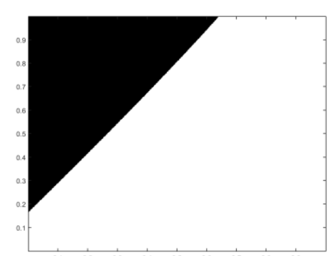
Equilibrium: 1.0, 0.75	**Buyer (—) and seller (···)**	**L1: 1.0, 0.75**
		
Equilibrium: 1.0, 1.0	**Buyer (—) and seller (···)**	**L1: 1.0, 1.0**
		
Equilibrium: 1.0, 1.25	**Buyer (—) and seller (···)**	**L1: 1.0, 1.25**
		
Equilibrium: 1.0, 1.5	**Buyer (—) and seller (···)**	**L1: 1.0, 1.5**
		
Equilibrium: 1.0, 1.75	**Buyer (—) and seller (···)**	**L1: 1.0, 1.75**
		

Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of *L1*-incentive-compatible mechanisms

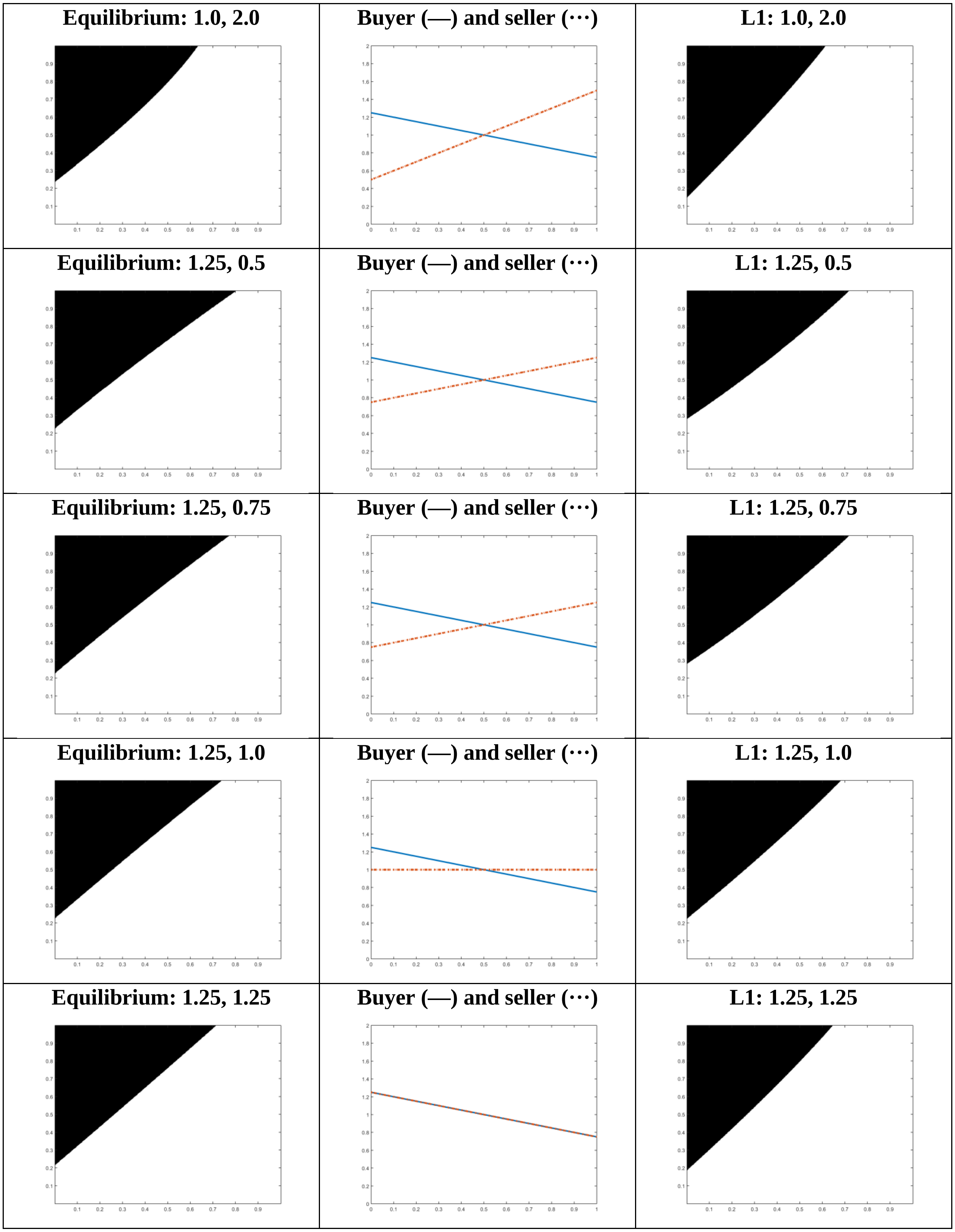


Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of *L1*-incentive-compatible mechanisms

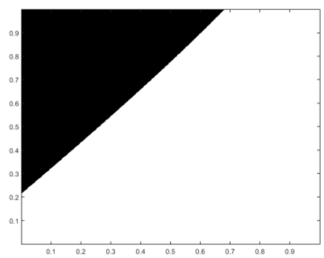
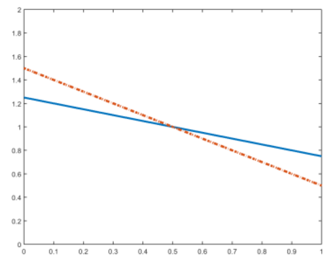
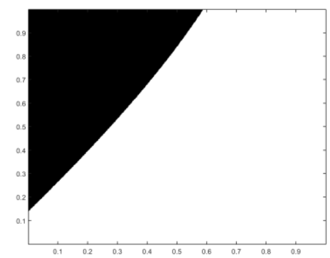
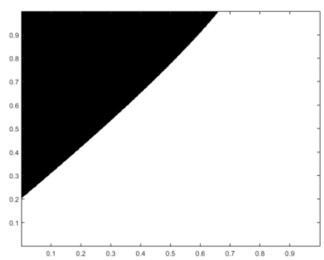
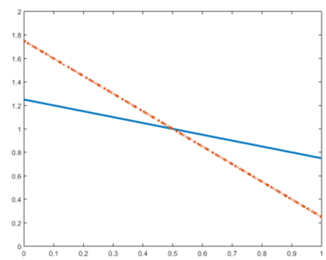
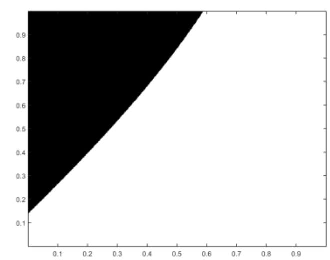
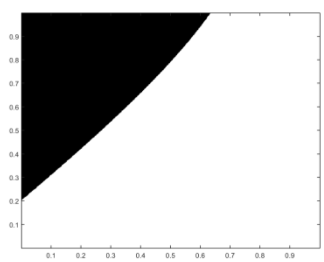
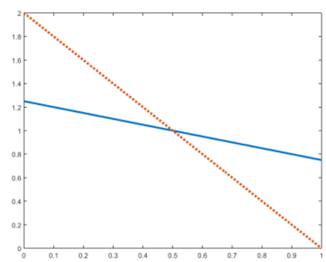
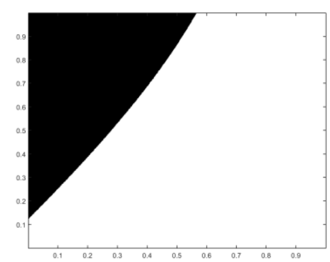
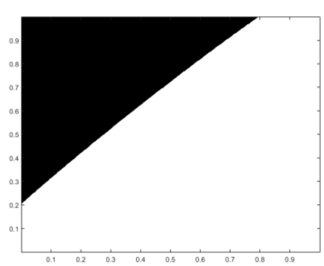
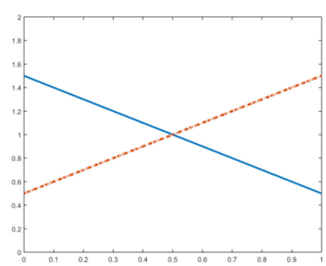
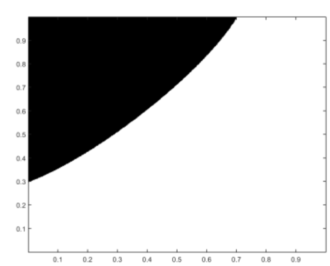
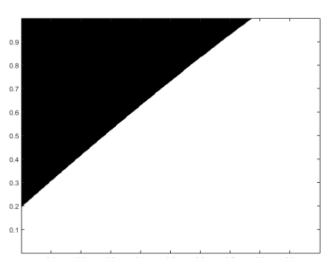
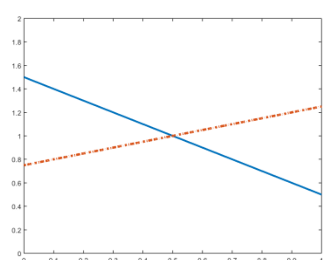
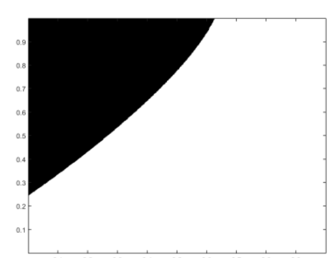
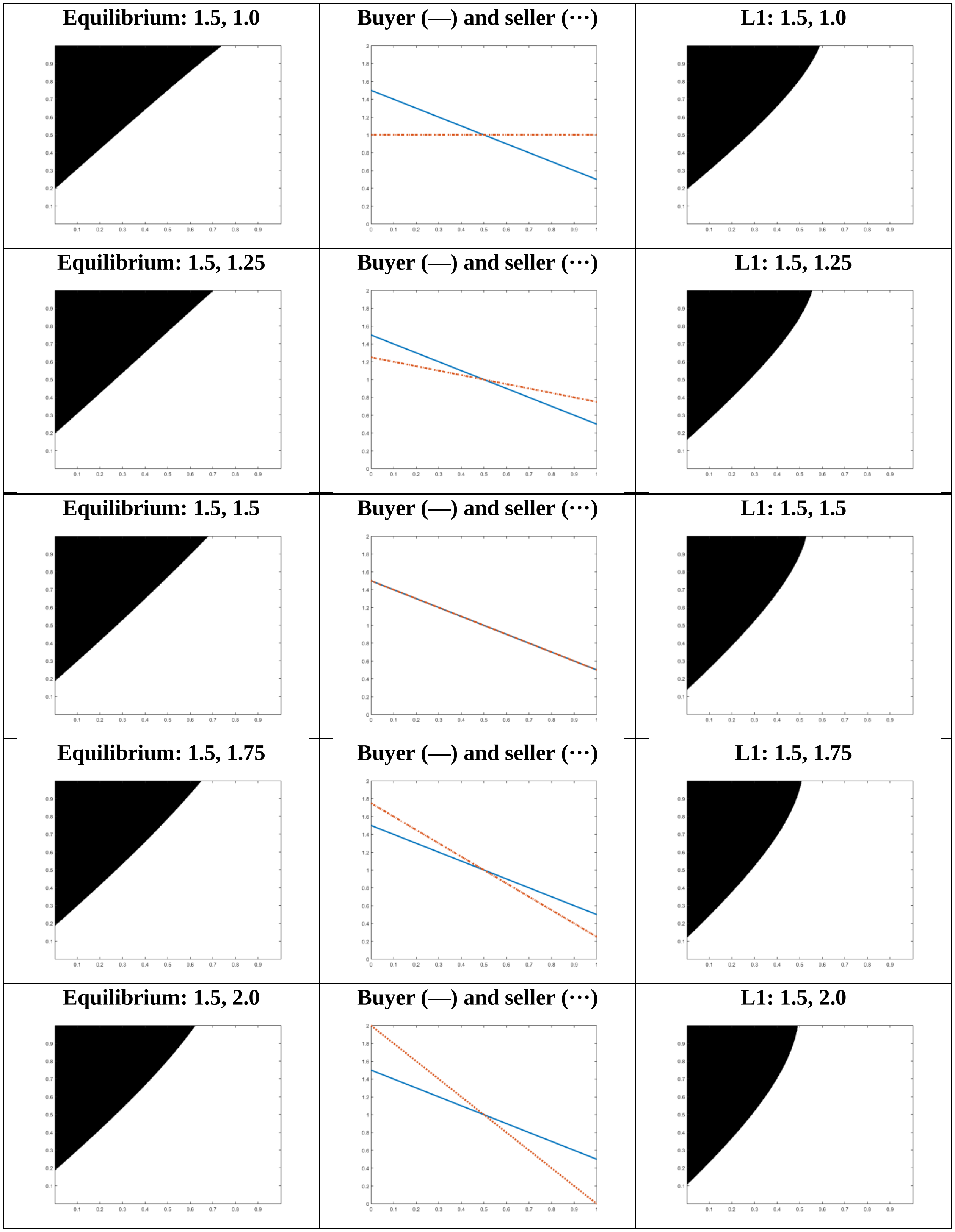
Equilibrium: 1.25, 1.5 	**Buyer (—) and seller (···)** 	**L1: 1.25, 1.5** 
Equilibrium: 1.25, 1.75 	**Buyer (—) and seller (···)** 	**L1: 1.25, 1.75** 
Equilibrium: 1.25, 2.0 	**Buyer (—) and seller (···)** 	**L1: 1.25, 2.0** 
Equilibrium: 1.5, 0.5 	**Buyer (—) and seller (···)** 	**L1: 1.5, 0.5** 
Equilibrium: 1.5, 0.75 	**Buyer (—) and seller (···)** 	**L1: 1.5, 0.75** 

Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of *L1*-incentive-compatible mechanisms



Online appendix: MATLAB code, developed by Rustu Duran, University of Oxford, for computation of mechanisms that are efficient in the set of *L1*-incentive-compatible mechanisms with a homogenous *L1* population, or that are equilibrium incentive-efficient.

The code assumes a homogeneous population of *L1*s and full-support linear densities of values *V* and *C*, each characterized by the value of *f*(0) or *g*(0), which, given linearity, range from 0 to 2, with *f*(0) = 1 or *g*(0) = 1 corresponding to uniform densities. *f*(0) and *g*(0) are represented by categorical index variables called fbar and gbar as follows. The interval [0, 2] is discretized into nine points, 0, 0.25, 0.50,…, 2.0, with index *i* representing the ith point. For instance, fbar = 2 means *f*(0)= 0.25 and gbar = 3 means *g*(0) = 0.5.

Solution algorithm

For *L1*s, the algorithm fixes a pair of value densities. For each given value of λ ("alfa" in the code) starting from 0.05, the code first uses the Kuhn-Tucker condition (5.14) to determine for which (v, c) combinations $p(v, c) = 1$. It then integrates the incentive budget constraint (5.5) (with $f^k(v; p, x) \equiv 1$ and $g^k(c; p, x) \equiv 1$ for *L1*s) for that λ. It then iterates these operations, increasing λ by increments of 0.05, until it finds the λ that makes the value on the right-hand side of (5.5) smallest; and checks how often that value changes sign (never more than once in the calculations). Finally, it chooses the value of λ that makes the value of the right-hand side as close to 0 as possible from above. This entire operation is done separately for each pair of value densities. Figure 1 is based on a coarse subset of all possible discretized combinations of linear value densities. A few extreme combinations are excluded because they violate the monotonicity conditions that, by Theorems 2 and 3, are needed for the mechanism to be truly optimal.

For equilibrium traders, the algorithm works in a completely analogous fashion.

To implement the algorithm, first run the program main.m. surf(exanteprobtrade) then shows how the ex ante probability of trade varies with the indices fbar (on the left axis) and gbar (on the right). surf(expectedtotalsurplus) shows how the expected total surplus varies with fbar and gbar. blackandwhite(fbar, gbar, pi) shows the trading region, with the area where $p(v, c) = 1$ in black.

To do the analogous computations for equilibrium traders, add "s" to the end of the arguments; e.g. surf(exanteprobtrades) instead of surf(exanteprobtrade). sidebyside(fbar1,gbar1,pi,pis,fbar,gbar) shows both the equilibrium and *L1* trading regions. comparetradearea (fbar1,gbar1,fbar2,gbar2,pi) compares the two trade areas, with a value of 0 meaning no trade in either case; 1 (2) only in the second (first) case; and 3 in both cases.

blackandwhite.m

```
%this is the function for visualisizng trade zone in lk model
function blackandwhite= blackandwhite (fbar,gbar,pi)
figure;
minuspi=1-pi(:,:,fbar,gbar);
blackandwhite=imagesc(minuspi);
 set( gca,'XTick',0:50:450,'XTickLabel',{'0','0.1', '0.2','0.3', '0.4',...
     '0.5',  '0.6', '0.7', '0.8', '0.9',},'YTick',0:50:450,...
     'YTickLabel',{'1','0.9','0.8', '0.7','0.6', '0.5', '0.4',...
     '0.3', '0.2', '0.1'} );
colormap gray;
end
```

blackandwhites.m

```
%this is the function for visualising trade zone in eqm model
function blackandwhite= blackandwhites (fbar,gbar,pis)
figure;
minuspis=1-pis(:,:,fbar,gbar);
blackandwhite=imagesc(minuspis);
set( gca,'XTick',0:50:450,'XTickLabel',{'0','0.1', '0.2','0.3', '0.4',...
    '0.5',  '0.6', '0.7', '0.8', '0.9', '1'},'YTick',0:50:450,...
    'YTickLabel',{'1','0.9','0.8', '0.7','0.6', '0.5', '0.4',  '0.3',...
    '0.2', '0.1'} );
colormap gray;
end
```

comparetradearea.m

```
% this function compares the trading areas.
% if the value of the function is 3, then it is a common trading area.
% if 2, then the first one trades but not the second one.
% if 1, only the second one trades. if 0, nobody trades.
```

```
function comparetradeareas = comparetradearea( fbar1,gbar1,fbar2,gbar2,pi)
  first= pi(:,:,fbar1,gbar1);
  second= pi(:,:,fbar2,gbar2);
  result= first*2+second;
  comparetradeareas=mesh(result);
end
```

fcdf.m

```
% cdf of buyers valuation, characterised by fbar
function fcdfc= fcdf (fbar,v)
fcdfc=fbar*v+v^2*(1-fbar);
end
```

fpdf.m

```
% pdf of buyers valuation, characterised by fbar
function fpdfc= fpdf(fbar,v)
fpdfc= fbar+v*2*(1-fbar);
end
```

gcdf.m

```
% cdf of sellers valuation, characterised by gbar
function gcdfc= gcdf (gbar,c)
gcdfc=gbar*c+c^2*(1-gbar);
end
```

gpdf.m

```
% pdf of seller's valuation, characterised by gbar
function gpdfc= gpdf(gbar,c)
gpdfc= gbar+c*2*(1-gbar);
end
```

main.m

```
clear all;
tic;        % chronometer mini-code for the elapsed time - toc is the
            % second part; at the end of the document.
beg= 0.001;    % beginning value for the discretised value range of seller
            % and buyer
fin= 0.999;    % ending value for the discretised value range of seller
            % and buyer
incr= 0.002;   % increment value for the discretised value range of the
            % seller and buyer
            % this value of the increment creates intervals each of
            % which is 0.002 unit lenght.
charbeg= 0;  % beginning value for the discretised characterising value
            % range of linear distributions
charfin= 2;  % ending value for the discretised characterising value
            % range of linear distributions
charincr= 0.25; % increment value for the discretised characterising value
            % range of linear distributions
            % characterising values represent the y-intercept of pdf.
            % I use alfa in order to refer lambda in the paper.
alfabeg=0.05;   % beginning value for the discretised value range of lambda
alfafin=1;      % ending value for the discretised value range of lambda
alfaincr=0.05;  % increment value for the discretised value range of lambda
v=beg:incr:fin; % generation of dicretised values of buyer
c=beg:incr:fin; % generation of dicretised values of seller
fbar=charbeg:charincr:charfin; % generation of dicretised characterising
                         % values of buyer's value distribution
gbar=charbeg:charincr:charfin; % generation of dicretised characterising
                         % values of seller's value distribution
alfa=alfabeg:alfaincr:alfafin; % generation of discretised lambdas
sumvecc= zeros (size(fbar,2), size (gbar,2), size (alfa,2));
```

```
% the vector I have created for integration (incentive-budget constraint)
sumvekk= zeros (size(fbar,2), size (gbar,2), size (alfa,2));
% the vector I have created for integration (incentive-budget constraint)
% - for equilibrium counterpart
counter= size(fbar,2)*size(gbar,2)*size(alfa,2);
% I use counter in order to be able to monitor the duration of the progress
norm=(1/((size(v,2))^2)); % normalisation for integration
for fbars=1:size(fbar,2)
    for gbars=1: size(gbar,2)
        for alfas= 1:1:size(alfa,2)
                        counter=counter-1
            for vs=1:size(v,2)
                for cs=1:size(c,2)
pi=pfunction(vs,cs,fbars,gbars,alfas,alfa,fbar,gbar,v,c);
% p(v,c) in the notes
pieq=pfunctioneq(vs,cs,fbars,gbars,alfas,alfa,fbar,gbar,v,c);
% p(v,c) in the notes
fdist=fpdf(fbar(fbars),v(vs)); % value of pdf of v.
gdist=gpdf(gbar(gbars),c(cs)); % value of pdf of c.
fcum=fcdf(fbar(fbars),v(vs));  % value of cdf of f.
gcum=gcdf(gbar(gbars),c(cs));  % value of cdf of c.
phis=(v(vs)/gdist)-((1-fcum)/(fdist*gdist));
%phi function in the paper
phiss=v(vs)-((1-fcum)/fdist);
%phi function in the paper-for eqm counterpart
gammas=(c(cs)/fdist)+((gcum)/(fdist*gdist));
%gamma function in the paper.
gammass=c(cs)+((gcum)/gdist);
%gamma function in the paper.-for eqm counterpart
sumvecc (fbars,gbars,alfas)=sumvecc(fbars,gbars,alfas)...
                      +(2*v(vs)-2*c(cs)-1)*pi*norm ;
```

```
                % integration
sumvekk (fbars,gbars,alfas)=sumvekk(fbars,gbars,alfas)...
                +(phiss-gammass)*pieq*fdist*gdist*norm ;
                % integration-for eqm counterpart
         end
       end
     end
   end
 end
% the following loop is in order to see how many times the integration
 % (as a function of lambda) intersects with the horizontal axis.
maximand=zeros(size(fbar,2), size (gbar,2));
for fbars=1:size(fbar,2)
   for gbars=1: size(gbar,2)
      for alfas= 2:1:size(alfa,2)
      if sumvecc(fbars,gbars,alfas-1)>0 && sumvecc(fbars,gbars,alfas)<0
        maximand(fbars,gbars)=maximand(fbars,gbars)+1;
      end
      if sumvecc(fbars,gbars,alfas-1)<0 && sumvecc(fbars,gbars,alfas)>0
        maximand(fbars,gbars)=maximand(fbars,gbars)+1;
      end
      end
   end
 end
% now we find the lambda which makes the integration closest to zero
 % for each fbar and gbar.
[minimisedvalues, indicesofbestalfas]=min(abs(sumvecc),[],3);
[minimisedvaluess, indicesofbestalfass]=min(abs(sumvekk),[],3);
 % this following step generates the pi matrices, which will be employed
 % for 2-dimensional graphs for trading regions for each fbar&gbar.
```

```
 pi=zeros(size(v,2),size(c,2),size(fbar,2),size(gbar,2));
pis=zeros(size(v,2),size(c,2),size(fbar,2),size(gbar,2));
 % for eqm counterpart
for fbars=1:size(fbar,2)
    for gbars=1: size(gbar,2)
        for vs=1:size(v,2)
             for cs=1:size(c,2)
  % for lk thinking
  bestalfa=indicesofbestalfas(fbars,gbars);
  pi(size(v,2)-vs+1,cs,fbars,gbars)=...
  pfunction(vs,cs,fbars,gbars,bestalfa,alfa,fbar,gbar,v,c);
  % and for eqm counterpart
  bestalfas=indicesofbestalfass(fbars,gbars);
  pis(size(v,2)-vs+1,cs,fbars,gbars)=...
  pfunctioneq(vs,cs,fbars,gbars,bestalfas,alfa,fbar,gbar,v,c);
             end
        end
    end
 end
% now we find ex-ante probability of trade and expected total surplus for
% each binary of fbar and gbar; for lk model
exanteprobtrade=zeros(size(fbar,2),size(gbar,2));
expectedtotalsurplus=zeros(size(fbar,2),size(gbar,2));
                    % and for eqm.
exanteprobtrades=zeros(size(fbar,2),size(gbar,2));
expectedtotalsurpluss=zeros(size(fbar,2),size(gbar,2));
for fbars=1:size(fbar,2)
    for gbars=1: size(gbar,2)
        for vs=1:size(v,2)
             for cs=1:size(c,2)
bestalfa=indicesofbestalfas(fbars,gbars);
```

```
bestalfas=indicesofbestalfass(fbars,gbars);
ppi=pfunction(vs,cs,fbars,gbars,bestalfa,alfa,fbar,gbar,v,c);
% p(v,c) in the paper.
ppis=pfunctioneq(vs,cs,fbars,gbars,bestalfas,alfa,fbar,gbar,v,c);
% p(v,c) in the paper.
fdist=fpdf(fbar(fbars),v(vs)); % value of pdf of v.
gdist=gpdf(gbar(gbars),c(cs)); % value of pdf of c.
exanteprobtrade(fbars,gbars)=...
   exanteprobtrade(fbars,gbars)+fdist*gdist*ppi*norm;
expectedtotalsurplus(fbars,gbars)=...
   expectedtotalsurplus(fbars,gbars)+ fdist*gdist*ppi*(v(vs)-c(cs))*norm;
exanteprobtrades(fbars,gbars)=...
   exanteprobtrades(fbars,gbars)+fdist*gdist*ppis*norm;
expectedtotalsurpluss(fbars,gbars)=...
   expectedtotalsurpluss(fbars,gbars)+ fdist*gdist*ppis*(v(vs)-c(cs))*norm;
                end
          end
     end
end
% this last piece of code is for saving trading regions,
% distribution functions and publishing the code
for fbars=1:size(fbar,2)
     for gbars=1: size(gbar,2)
         ef=100*fbar(fbars);
         gi=100*gbar(gbars);
         name1= num2str(ef);
         name2= num2str(gi);
         namel1 = strcat(name1,name2,'L1');
         nameeqm = strcat(name1,name2,'eqm');
         l1= blackandwhite(fbars,gbars,pi);
         eqm=blackandwhites(fbars,gbars,pis);
```

```
        saveas(l1,namel1,'png');
        saveas(eqm,nameeqm,'png');
        yeni (fbars,gbars,fbar,gbar)
    end
end
toc;
```

pfunction.m

```
% this is the value of a p(v,c), in lk model, for particular values of
% v,c,fbar,gbar,alfa
function pfuncc=pfunction (vs,cs,fbars,gbars,alfas,alfa,fbar,gbar,v,c)
fdist=fbar(fbars)+v(vs)*2*(1-fbar(fbars));        % pdf of v.
gdist=gbar(gbars)+c(cs)*2*(1-gbar(gbars));        % pdf of c.
fcum=fbar(fbars)*v(vs)+v(vs)^2*(1-fbar(fbars));    % cdf of f.
gcum=gbar(gbars)*c(cs)+c(cs)^2*(1-gbar(gbars));    % cdf of c.
            x=  v(vs)-c(cs) ;
            y=(1-fcum)/(fdist*gdist);
            z= gcum/(fdist*gdist);
            r= x-(alfa(alfas))*(y+z);
              if r>=0
                  pfuncc=1;
              else
                  pfuncc=0;
              end
end
```

pfunctioneq.m

```
% this is the value of a p(v,c), in usual model, for particular values of
% v,c,fbar,gbar,alfa
```

```
function pfuncc=pfunctioneq (vs,cs,fbars,gbars,alfas,alfa,fbar,gbar,v,c)
fdist=fbar(fbars)+v(vs)*2*(1-fbar(fbars));        % pdf of v.
gdist=gbar(gbars)+c(cs)*2*(1-gbar(gbars));        % pdf of c.
fcum=fbar(fbars)*v(vs)+v(vs)^2*(1-fbar(fbars));    % cdf of f.
gcum=gbar(gbars)*c(cs)+c(cs)^2*(1-gbar(gbars));    % cdf of c.
            x=  v(vs)-c(cs) ;
            y= (v(vs))-((1-fcum)/(fdist));
            z= (c(cs))+(gcum/(gdist));
            r= x+(alfa(alfas))*(y-z);
              if r>=0
                 pfuncc=1;
              else
                 pfuncc=0;
              end
end
```

sidebyside.m

```
%this is the function for visualisizng trade zone in lk model side by side
function [blackandwhite middle blackandwhites]=...
   sidebyside (fbars,gbars,pi, pis,fbar,gbar)
figure;
minuspi=1-pi(:,:,fbars,gbars);
minuspis=1-pis(:,:,fbars,gbars);
subplot(1,3,1);
blackandwhite=imagesc(minuspi);
title('trade zone in L1 model');
 set( gca,'XTick',0:50:450,'XTickLabel',{'0','0.1', '0.2','0.3', '0.4',...
     '0.5',  '0.6', '0.7', '0.8', '0.9',},'YTick',0:50:450,'YTickLabel',...
     {'1','0.9','0.8', '0.7','0.6', '0.5', '0.4',  '0.3', '0.2', '0.1'} );
axis image;
subplot(1,3,3);
```

```
blackandwhites=imagesc(minuspis);
 set( gca,'XTick',0:50:450,'XTickLabel',{'0','0.1', '0.2','0.3', '0.4',...
     '0.5',  '0.6', '0.7', '0.8', '0.9',},'YTick',0:50:450,'YTickLabel',...
     {'1','0.9','0.8', '0.7','0.6', '0.5', '0.4',  '0.3', '0.2', '0.1'} );
title('trade zone in equilibrium');
axis image;
subplot(1,3,2);
i = linspace(0,1);
buyer= fbar(fbars)+i*2*(1-fbar(fbars));
seller= gbar(gbars)+i*2*(1-gbar(gbars));
middle=plot(i,buyer,'-',i,seller,':');
title(' buyer (-) and seller (..) ');
axis image;
colormap gray;
end
```

yeni.m

```
%this is the function for visualising distributions
function [middle]= yeni (fbars,gbars,fbar,gbar)
h=figure;
        ef=100*fbar(fbars);
        gi=100*gbar(gbars);
        name1= num2str(ef);
        name2= num2str(gi);
        f=fbar(fbars);
        g=gbar(gbars);
        namedensity=strcat(name1,name2,'density');
i = linspace(0,1);
buyer= f+i*2*(1-f);
seller= g+i*2*(1-g);
middle=plot(i,buyer,'-',i,seller,':','linewidth',3);
```

```
title(' ');
ylim([0,2]);
colormap gray;
saveas(h,namedensity,'png');
end
```