

Efficient Mechanisms for Level- k Bilateral Trading

Vincent P. Crawford¹

First version 18 April 2015; this revision 28 May 2019

Abstract: I revisit Roger Myerson and Mark Satterthwaite's (1983; "MS") analysis of mechanism design for bilateral trading, replacing equilibrium with a level- k model and focusing on direct mechanisms. The revelation principle fails for level- k models. However, if only level- k -incentive-compatible mechanisms are feasible and traders' levels are observable, MS's characterization of mechanisms that maximize traders' expected gains from trade subject to incentive constraints generalizes, with one novel feature. If traders' levels are unobservable, only random posted-price mechanisms are level- k -incentive-compatible, and a particular deterministic posted-price mechanism maximizes expected gains from trade. If non-level- k -incentive-compatible mechanisms are feasible, optimal mechanisms may differ more extensively.

Keywords: mechanism design, bilateral trading, level- k thinking

JEL codes: C70, D02

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. A preliminary version of this paper was presented as the Economic Journal Lecture at the Royal Economic Society Conference, London, April 2011. The first complete version was dated 18 April 2015. This version corrects significant errors in the proofs of Theorems 2 and 3 and the claimed proof of Corollary 1. Special thanks are due to Rene Saran for his careful reading and valuable advice. I also owe thanks to Johannes Abeler, B. Duoglas Bernheim, Colin Camerer, Yan Chen, Miguel Costa-Gomes, Jacques Crémer, Avinash Dixit, Kfir Eliaz, Tore Ellingsen, Xavier Gabaix, Olga Gorelkina, Nagore Iriberry, Paul Klemperer, Margaret Meyer, Paul Milgrom, Roger Myerson, Philip Neary, Zvika Neeman, Nick Netzer, Robert Östling, Ady Pauzner, Phil Reny, Larry Samuelson, Adam Sanjurjo, Roberto Serrano, Joel Sobel, Alex Teytelboym, and anonymous 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-BESTDECISION. 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 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, which evidence suggests better predicts initial responses to games. I also focus on direct mechanisms. Otherwise I maintain standard assumptions about behavior and design.

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? A design may still need to work the first time; and design creates new games, which may lack the clear precedents required for learning. Further, even if learning is possible, a design may create games too complex for convergence to equilibrium to be behaviorally plausible.

Even without learning, the equilibrium assumption can still be logically 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 subjects' thinking often favors level- k decision rules, which anchor beliefs in a naive model of others' initial responses called $L0$, usually taken to be to be uniformly random over the feasible decisions; and then adjust them via a small number (k) of iterated best responses: $L1$ best responds to $L0$, $L2$ to $L1$, and so on. The estimated frequency of $L0$ is usually zero or very small; and the estimated distribution of subjects' levels is normally concentrated on the lowest two or three levels.

Lk for $k > 0$ is decision-theoretically rational, with an accurate model of the game. It departs from equilibrium only in basing beliefs on an oversimplified model of others. Lk 's decisions respect k -rationalizability (Bernheim 1984), so it mimics equilibrium decisions in two-person games that are dominance-solvable in k rounds, but can deviate systematically in other games.³

² Crawford, Costa-Gomes, and Iriberry (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 are likely 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 $k > 1$.

Importantly, a structural model based on level- k thinking not only predicts that deviations from equilibrium will sometimes occur, but also which kinds of game evoke them and what forms they will take. It also replaces k -rationalizability's set-valued predictions with a specific selection, which permits an analysis with precision close to that of an equilibrium analysis.⁴

A level- k analysis of design can yield several benefits. It can clarify the role of equilibrium assumptions in analyses like MS's. It can identify settings where equilibrium-based conclusions are robust to likely deviations from equilibrium; and others where mechanisms that are optimal if equilibrium is assumed may perform worse in practice than others that are more robust. Finally, a level- k analysis might reduce the sensitivity of incentive-efficient mechanisms to distributional and knowledge assumptions that institutions in use seldom respond to (Robert Wilson 1987).

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

MS asked whether the ex post inefficiency in CS's equilibria is an avoidable flaw of the double auction, or rather a general property of any trading mechanism that must create incentives for traders to (explicitly or implicitly) reveal their private values, as efficient trading requires.

Section II reviews MS's equilibrium-based analysis of design for CS's setting. MS's analysis began with the revelation principle (pp. 267-268), the observation that any given equilibrium of the game created by any feasible mechanism can be viewed as the truthful equilibrium of some direct-revelation mechanism. Thus, assuming equilibrium and a given selection, 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. For well-behaved value densities with overlapping supports, MS then characterized the mechanism that maximizes traders' expected

⁴ Until recently the alternatives to assuming equilibrium were largely 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 precision allows an analysis that yields additional insight. A cognitive hierarchy version of my analysis would also be feasible, but would involve some loss of clarity.

total gains from trade in the set of incentive-compatible and interim individually rational mechanisms.⁵ MS then used their characterization to show that, in equilibrium, no feasible mechanism can avoid the ex post inefficiency of CS's double auction. MS also showed that with uniform densities, CS's linear double-auction equilibrium, or equivalently, via the revelation principle, the incentive-compatible mechanism that mimics its outcomes in equilibrium, maximizes traders' expected total gains from trade subject to the incentive constraints.

Section III defines a level- k model for incomplete-information games. I restrict attention throughout to games created by direct mechanisms, in which players' decisions can be viewed as estimates of their values. This involves significant loss of generality, but only for direct games is there clear experimental evidence to guide the specification, and the simplicity of such games is an important advantage in applications. Following previous work (Crawford et al. 2013, Section 5), I generalize the usual complete-information specification that LO 's decisions are uniformly random over the feasible decisions by taking LO 's decisions to be uniform over those decisions and independent of its own private value. I define $L1, L2, \dots$ via iterated best responses as usual. The resulting model gives a reliable account of how people's thinking deviates from equilibrium and of their "informational naiveté"—the imperfect attention to how others' decisions depend on their private information that is often observed in phenomena like the winner's curse.⁶

In preparation for Section V's analysis of level- k design, Section IV extends CS's equilibrium analysis of the double auction to Section III's generalized level- k model. I focus on $L1$ s and $L2$ s, which predominate empirically and illustrate the main points. With uniform value densities, $L1$ s ask or bid systematically more aggressively than in equilibrium; and their aggressiveness drives their expected total gains from trade when matched well below its level in CS's linear equilibrium. By contrast, with uniform densities $L2$ s bid or ask less aggressively than in CS's equilibrium; that raises their gains from trade when matched above its equilibrium level.

Section V begins the level- k design analysis. When needed for clarity, I use the prefixes "equilibrium" and "level- k " to distinguish standard concepts like incentive-compatibility and interim individual rationality that depend on a trader's beliefs.

⁵ Steven Williams (1987) notes that MS's maximization of the expected gains from trade does not identify all mechanisms associated with outcomes on the incentive-efficient frontier, because the incentive constraints interfere with the transferable utility that usually follows from quasilinear utilities. He characterizes the mechanisms associated with other welfare weights.

⁶ With independent private values, the winner's curse is not relevant here; but other kinds of informational naiveté are relevant.

Theorem 1 records the fact that, with uniform value densities, the sets of level- k -incentive-compatible and interim individually rational mechanisms and equilibrium-incentive-compatible and interim individually rational mechanisms coincide for any population of level- k traders with $k > 0$. In this case, MS's characterization of the mechanism that maximizes traders' expected total gains from trade in the set of incentive-compatible and interim individually rational mechanisms generalizes completely to the level- k model, with one qualification: Comparing Theorem 1 with Section IV's level- k analyses of the double auction reveals an important limitation of the revelation principle for level- k design. For equilibrium traders with uniform value densities, MS's mechanism that maximizes traders' expected gains from trade in the set of equilibrium-incentive-compatible and interim individually rational mechanisms is outcome-equivalent to the double auction. But for $L1$ traders, the double auction then yields lower gains from trade than MS's mechanism, while for $L2$ traders it yields higher gains. For level- k traders, MS's mechanism maximizes expected total gains from trade subject to the level- k incentive constraints only if implemented in its equilibrium- and level- k -incentive-compatible form.

The revelation principle does not work as expected here because the choice between mechanisms that are outcome-equivalent in equilibrium influences the correctness of level- k beliefs by changing the mechanism's relationship to $L0$, via Crawford, Tamar Kugler, Zvika Neeman, and Ady Pauzner's (2009; "CKNP") "level- k menu effects". Replacing the double auction with MS's equilibrium- and level- k -incentive-compatible mechanism rectifies $L1$ and $L2$ beliefs, increasing gains from trade to their equilibrium level for $L1$ s but reducing them for $L2$ s.

Because requiring level- k -incentive-compatibility involves loss of generality, in applications one must decide whether incentive-compatibility is truly necessary, or can be relaxed to allow non-incentive-compatible direct mechanisms like the double auction. Section V considers design when level- k -incentive-compatibility is required and Section VI considers relaxing it.

In each case the results depend on whether traders' levels are observable or predictable (henceforth shortened to "observable"). When traders' levels are observable, I assume that the mechanism can be tailored to their combination of levels. In that case, when level- k -incentive-compatibility and interim individual rationality are required, Theorems 2-3 show that MS's characterization of mechanisms that maximizes traders' expected total gains from trade for general well-behaved value densities generalizes, qualitatively, to the level- k model. Design features that enhance a mechanism's performance in MS's analysis also enhance it in the level- k

analysis, but with different weights. The level- k analysis identifies another, novel feature that enhances performance, which I call tacit exploitation of predictably incorrect beliefs (“TEPIB”). MS’s Corollary 1, that no incentive-compatible mechanism can assure ex post efficient trade with probability one, does *not* generalize, but level- k incentives still limit ex post efficient trade.

Theorem 4 turns to the case where level- k -incentive-compatibility and interim individual rationality are required but traders’ levels are unobservable, so that the mechanism must screen traders’ levels as well as their values. The level- k incentive constraints then compel the use of a (possibly random) posted-price mechanism; and a mechanism that maximizes traders’ expected gains from trade subject to the level- k incentive constraints must be equivalent to a particular deterministic posted-price mechanism. Such a mechanism makes truthful bidding a weakly dominant strategy for all levels greater than 0, and thus reconciles the level- k and equilibrium incentive constraints. The optimal posted price is independent of the distribution of levels, but it depends on traders’ value densities. It can however be implemented via a dynamic process that is independent of those densities (Jernej Čopič and Clara Ponsatí 2008, 2016), thereby satisfying Wilson’s (1987) desideratum and bringing the optimal mechanism closer to those used in practice. The posted-price mechanism cannot assure ex post efficient trade with probability one; but at least in the case of uniform value densities, the cost of its limited sensitivity is modest.

These results elucidate the rationale for robust mechanism design. Theorems 2-3 show that nonequilibrium design when traders’ levels are observable involves more than implementing the best equilibrium outcome under weaker behavioral assumptions.⁷ Theorem 4 derives the need for implementation in weakly dominant strategies, often simply assumed in robust mechanism design, from the unpredictability of strategic thinking, bringing its rationale closer to intuition.

I conduct the formal analysis for level- k models because they are well supported by evidence, and for concreteness. But Section V’s analysis of the cases where level- k -incentive-compatibility is required goes through, qualitatively, for any model that respects decision-theoretic rationality and makes generically unique predictions well-behaved enough to satisfy monotonicity restrictions like MS’s. Put another way, MS’s equilibrium assumption bundles decision-theoretic rationality with homogeneity and statistical correctness of traders’ beliefs, and the revelation principle sidesteps the issue of uniqueness. My analysis shows that one can largely dispense with

⁷ CKNP illustrate this point for auction design, showing that revenue-equivalence fails for level- k bidders, and that although a second-price auction yields the equilibrium-optimal revenue level for level- k as well as for equilibrium bidders, a first-price auction yields more revenue when LJ bidders are common.

homogeneity, statistical correctness, and uniqueness, which are behaviorally the least plausible of MS's assumptions and which at first sight appear to play essential roles in their analysis.

Continuing to focus on direct mechanisms, Section VI records some observations about design when level- k -incentive-compatibility is not required.⁸ Then, a direct mechanism's effects cannot be tractably captured via incentive constraints, and must instead be modeled directly. I show that whether or not traders' levels are observable, when level- k -incentive-compatibility is not truly necessary, relaxing it can yield optimal mechanisms that differ qualitatively from those that are optimal assuming equilibrium, possibly with significantly larger gains from trade.

Section VII is a brief conclusion.

In addition to CS's and MS's analyses, this paper builds on Crawford and Iriberri's (2007) level- k analysis of sealed-bid auctions and CKNP's level- k analysis of optimal independent-private-value auctions, which builds on Myerson's (1981) equilibrium analysis of auction design.

This paper's closest relatives in the recent literature are Geoffroy de Clippel, Saran, and Roberto Serrano (2019; see also 2016) and Terri Kneeland (2018). They both study level- k implementation of social choice rules, allowing general distributions of unobservable levels, and indirect mechanisms in which players report their levels as well as their private information.

Kneeland assumes uniform random LO s and allows mechanisms to treat levels unequally, as here; and considers both single- and set-valued rules. She proves general necessary and sufficient conditions for level- k implementation that is robust to variations in players' beliefs about others' values and levels, which amounts to requiring ex post incentive-compatibility. For single-valued rules or direct mechanisms, such robustness makes level- k and equilibrium incentive constraints coincide. In general, however, she shows that robust level- k incentive constraints are weaker than equilibrium incentive constraints. As a result, in an environment near MS's, there are set-valued indirect mechanisms, in which players' report their levels as well as their values and which may treat levels unequally, that robustly assure ex post efficient trade with probability one—in contrast to my result for level- k incentive-compatible direct mechanisms in MS's setting.

De Clippel et al. require implementation independent of LO s within a large class, and equal treatment of levels.⁹ Like Kneeland, they show under mild restrictions that for single-valued rules, robust level- k and equilibrium incentive constraints coincide. Requiring single-valued rules

⁸ That is, allowing direct mechanisms that create incentives to lie, but continuing to assume that traders best respond.

⁹ De Clippel et al. argue that treating levels equally is standard, but there are few if any precedents for how to treat levels of reasoning, and in mainstream screening analyses it is seldom assumed that private-information types must be treated equally.

and equal treatment of levels, they reach the opposite conclusion from Kneeland's about whether ex post efficient trade with probability one can be assured with level- k traders in MS's setting.

De Clippel et al.'s and Kneeland's results allowing heterogeneous, unobservable levels closely parallel my Theorem 4, which shows in MS's setting that when level- k incentive-compatibility is required, screening both values and levels requires a (possibly random) posted-price mechanism, and that a particular deterministic posted-price mechanism is optimal. My result would hold for De Clippel et al.'s wider set of LO anchors, and would make their equal treatment of levels a conclusion rather than an assumption. As already noted, Theorem 4 derives the need for robustness from the unobservability or unpredictability of strategic thinking, rather than simply assuming it. However, unlike Kneeland's and De Clippel et al.'s results for this case, mine are specific to MS's bilateral trading environment. But their analyses have no counterpart to my results in Theorems 2-3 and Section VI for the case where traders' levels are observable or approximately predictable, which I view as complementary and equally informative.¹⁰

In other work on nonequilibrium design, Hagerty and Rogerson (1987), Jeremy Bulow and John Roberts (1989, relaxing ex post budget balance), and Čopič and Ponsatí (2008, 2016) study dominant-strategy or distribution-free implementation in MS's setting. Saran (2011a) studies MS's design problem when some traders report truthfully without regard to incentives; Saran (2011b) studies how menu-dependent preferences affect the revelation principle; and Saran (2016) studies implementation with complete information when players' levels of rationality are heterogeneous and bounded, obtaining a version of the revelation principle. Tilman Börgers and Jiangtao Li (in press) characterize mechanisms that are "strategically simple" in the sense that players need form only first-order beliefs, with applications to voting and bilateral trade. Olga Gorelkina (2018) conducts a level- k analysis of the expected-externality mechanism, and 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 and Stephen Morris (2009) and Bergemann, Morris, and Olivier Tercieux (2011) study implementation in rationalizable strategies.¹¹

¹⁰ Experimental evidence (e.g. Marina Agranov, Elizabeth Potamites, Andrew Schotter, and Chloe Tergiman 2012; David Gill and Victoria Prowse 2016) suggests that subjects' levels are often correlated with observables, and such considerations sometimes influence applications (e.g. Parag Pathak 2017, Section 2.4.2).

¹¹ Kobi Glazer and Ariel Rubinstein (1998), Neeman (2003), Kfir Eliaz and Ran Spiegler (2006, 2007, 2008), and Alexander Wolitzky (2016) study design when the "behavioral" aspect concerns individual decisions or judgment. Björn Bartling and

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 $\frac{1}{2}$. Then, if $b \geq a$, the buyer acquires the object at price $(a + b)/2$, the seller's utility is $(a + b)/2 - C$, and the buyer's is $V - (a + b)/2$. If $b < a$, the seller retains the object, no money changes hands, and seller's and buyer's utilities are both 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 $b \in [0, 1]$

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

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 $a \in [0, 1]$

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

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

Nick Netzer (2016) and Felix Bierbrauer and Netzer (2016) study robustness to various kinds of social preferences in auction design and implementation of social choice rules.

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. But MS showed that in this case the double auction maximizes traders' expected total gains from trade subject to incentive constraints: assuming equilibrium, no mechanism can Pareto-dominate the linear equilibrium of the double-auction. That equilibrium yields ex ante probability of trade $9/32 \approx 28\%$ and expected total gains from trade $9/64 \approx 0.14$, less than the maximum ex post individually rational probability of trade of 50% and expected gains from trade $1/6 \approx 0.17$.

II. EQUILIBRIUM MECHANISM DESIGN FOR BILATERAL TRADING

Assuming equilibrium, MS characterized mechanisms that maximize traders' ex ante total expected gains from trade in CS's trading environment, allowing any feasible mechanism and taking into account the need to ensure equilibrium-incentive-compatibility and equilibrium-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.¹² For a direct mechanism $p(\cdot, \cdot)$ 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):

$$(2.1) \quad \begin{aligned} X_B(v) &= \int_0^1 x(v, \hat{c})g(\hat{c})d\hat{c}, & X_S(c) &= \int_0^1 x(\hat{v}, c)f(\hat{v})d\hat{v}, \\ P_B(v) &= \int_0^1 p(v, \hat{c})g(\hat{c})d\hat{c}, & P_S(c) &= \int_0^1 p(\hat{v}, c)f(\hat{v})d\hat{v}, \\ U_B(V, v) &= VP_B(v) - X_B(v), & U_S(C, c) &= X_S(c) - CP_S(c). \end{aligned}$$

Although the outcome functions take only 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 $[0, 1]$,

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

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

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

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

$$(2.4) \quad \begin{aligned} U_B(0,0) + U_S(1,1) &= \min_{V \in [0,1]} U_B(V, V) + \min_{C \in [0,1]} U_S(C, C) \\ &= \int_0^1 \int_0^1 \left(\left[V - \frac{1 - F(V)}{f(V)} \right] - \left[C + \frac{G(C)}{g(C)} \right] \right) p(V, C)g(C)f(V)dCdV. \end{aligned}$$

¹² Thus, MS assume the mechanism satisfies ex post (expected) budget balance, as I will assume in Section V's level- k analysis.

Furthermore, if $p(\cdot, \cdot)$ is any function mapping $[0, 1] \times [0, 1]$ 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

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

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:

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

(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

$$(2.7) \quad \int_0^1 \int_0^1 (V - C) p(V, C) g(C) f(V) dC dV \\ = U_B(0, 0) + U_S(1, 1) + \int_0^1 \int_0^1 [G(C) f(V) + \{1 - F(V)\} g(C)] p(V, C) dC dV.$$

(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 at (5.9) 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 (recall the limitation noted in footnote 5). Define

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

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 total 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 total 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. But the buyer may pay even if he does not get the object, violating ex post individual rationality (MS, p. 271).

With uniform value densities, MS's Theorem 2 allows a closed-form solution for the incentive-compatible, interim individually rational mechanism that maximizes total 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

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

and (2.4) with equality reduces to

$$(2.10) \quad 0 = \int_{\frac{c}{1+\alpha}}^1 \int_0^V \frac{1}{1+\alpha} (2V - 1 - 2C) dC dV = \frac{3\alpha-1}{6(1+\alpha)^3},$$

which implies that $\alpha = 1/3$ (MS, p. 277). The total-surplus-maximizing 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 total-surplus-maximizing mechanism.¹³ This yields ex ante probability of trade $9/32 \approx 28\%$ and expected gains from trade $9/64 \approx 0.14$.

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; but many of my results also hold for any nonequilibrium model that makes generically unique, well-behaved predictions. I focus on games like those created by direct mechanisms, in which players' decisions are conformable to value estimates, because their simplicity makes them especially well-suited to applications and evidence is lacking to guide a specification for more complex, general games.

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 Iriberry (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*.¹⁴ My analysis goes through if the

¹³ However, Satterthwaite and Steven Williams (1989, Theorem 5.1) showed that for generic value densities CS's double auction equilibria are incentive-*inefficient*. Thus MS's remarkable result for the case of uniform value densities is a coincidence.

¹⁴ Milgrom and Stokey's (1982) notions of Naïve Behavior and First-Order Sophistication, which they suggested might explain zero-sum trades despite their equilibrium-based No Trade or "Groucho Marx" Theorem, are equivalent to an $L1$ defined this way and an $L2$ best responding to such an $L1$. 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 would take $L0$ 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, and it would trivially reduce level- k incentive constraints to the equilibrium incentive constraints.

value densities have overlapping supports, provided that $L0$ is anchored 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 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 $k = 1$ and 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 $b \in [0, 1]$

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

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.¹⁵

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

$$\int_a^1 \left[\frac{a+b}{2} - C \right] db.$$

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', but are $1/12$ more aggressive. When an *L1* buyer meets an *L1* seller, trade takes place whenever $V \geq C + 1/2$, so the value gap needed for trade is $1/4$ larger than for equilibrium traders and ex post efficiency is lost for more values. An *L1* buyer's and seller's ex ante probability of trade is $1/8 = 12.5\%$ and their expected gains from trade is $1/24 \approx 0.04$, far less than the equilibrium probability $9/32 \approx 28\%$ and expected gains from trade $9/64 \approx 0.14$, and further below the maximum individually rational probability 50% and expected gains $1/6 \approx 0.17$.

B. *L2* traders

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

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

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 $b \in [1/3, 1]$

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

¹⁵ Compare Crawford and Iriberrí'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 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 $(3/2)(V - b) - (3/4)(V - 1/3) = 0$ 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 $a \in [0, 1]$

$$\int_a^1 \left[\frac{a+b}{2} - C \right] f(b_1^{-1}(b)) db,$$

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 $a \in [0, 2/3]$

$$\int_a^{2/3} \left[\frac{a+b}{2} - C \right] (3/2) db.$$

The second-order condition is satisfied, and the first-order condition $(3/2)(a-C) + (3/2)(2/3 - C)/2 = 0$ 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, and ex post efficiency is lost for fewer values. The probability of trade is $25/72 \approx 35\%$ and expected gains from trade is $11/72 \approx 0.15$, higher than the equilibrium

probability $9/32 \approx 28\%$ and expected gains from trade $9/64 \approx 0.14$ (Section II.B), but still well below the maximum individually rational probability 50% and expected surplus $1/6 \approx 0.17$.¹⁶

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 use the level- k model because it is well supported by evidence, and for concreteness; but the proofs show that most of my results hold for any nonequilibrium model that makes unique predictions that respect decision-theoretic rationality, and which satisfy monotonicity restrictions like those needed for MS's Theorems 1-2. I restrict attention to direct mechanisms because for them there is clear evidence to guide the specification of a nonequilibrium behavioral model, and their simplicity is important in applications. Finally, as in MS's and almost all other analyses of design, I assume that traders' responses are noiseless.

I define level- k -incentive-compatibility and interim individual rationality as for the usual notions, but for level- k beliefs. Expected total gains from trade are defined for the correct beliefs.

A. Uniform value densities

First consider the case of uniform value densities, in which MS obtained a closed-form solution for the direct mechanism that maximizes traders' expected total gains from trade in the set of equilibrium-incentive-compatible mechanisms, which then mimics CS's linear double-auction equilibrium. Theorem 1 shows that with uniform value densities, MS's expected-gains-maximizing direct mechanism is fully robust to relaxing equilibrium in favor of a level- k model.

Theorem 1. *With uniform value densities, for any population of level- k traders with $k > 0$, whether or not traders' levels are observable, the set of level- k -incentive-compatible and interim individually rational mechanisms coincides with the set of equilibrium-incentive-compatible and interim individually rational mechanisms; and MS's direct mechanism that maximizes traders' expected total gains from trade in the set of equilibrium-incentive-compatible and interim individually rational mechanisms also maximizes traders' expected total gains from trade in the set of level- k -incentive-compatible and interim individually rational mechanisms.*

Proof. $L1$ traders believe they face a uniform distribution of the other trader's reports, so that with uniform value densities, $L1$'s incentive-compatibility and interim individual rationality

¹⁶ 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.

constraints ((5.2)-(5.3) and (5.5)-(5.6) below) coincide with those of equilibrium traders ((2.2)-(2.3) and (2.5)-(2.6), or MS's (2)-(4)). By induction, when $L1$ s report truthfully, $L2$ s' incentive-compatibility and individual rationality constraints coincide with those of equilibrium traders, and so on ad infinitum. Thus, MS's direct mechanism that maximizes traders' expected total gains from trade in the set of equilibrium-incentive-compatible mechanisms also maximizes traders' expected total gains from trade in the set of level- k -incentive-compatible and interim individually rational mechanisms for any population of level- k traders with $k > 0$.¹⁷ Q.E.D.

B. *Level- k menu effects and the revelation principle*

Comparing Theorem 1 with Section IV's level- k analyses of the double auction reveals a limitation of the revelation principle for level- k design. For equilibrium traders with uniform value densities, MS's direct mechanism that maximizes traders' ex ante expected total gains from trade in the set of equilibrium-incentive-compatible and interim individually rational mechanisms is outcome-equivalent to CS's linear double-auction equilibrium, in which traders shade bids and asks to mimic truthful reporting in MS's mechanism. But in Section IV's examples with uniform value densities, $L1$ s do worse in the double auction than in MS's mechanism, while $L2$ s do better. MS's mechanism neutralizes $L1$ s' aggressiveness in the double auction by rectifying their beliefs. MS's mechanism could also be used to neutralize $L2$ s' unaggressiveness, but unless level- k -incentive-compatibility is required, the non-incentive-compatible double auction preserves $L2$ s' non-equilibrium beliefs, allowing $L2$ traders to do better than in MS's mechanism. If level- k -incentive-compatibility is required, MS's mechanism maximizes total expected gains from trade only if implemented in its incentive-compatible form.

The revelation principle does not work as expected because the choice between mechanisms, even if outcome-equivalent in equilibrium, can change the mechanism's relationship to $L0$, influencing the correctness of level- k beliefs via CKNP's level- k menu effects. Thus the revelation principle may fail, and it matters whether level- k -incentive-compatibility is required or can be relaxed to allow traders to respond optimally to any direct mechanism.

Some 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 almost always in equilibrium analyses where there is no

¹⁷ 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 $L0$.

theoretical gain from relaxing it. Other analysts are willing to consider mechanisms that are not equilibrium-incentive-compatible, such as the first-price sealed-bid auction (Myerson 1981) or the Boston mechanism in school choice (Aytek Erdil and Haluk Ergin 2008; Abdulkadiroglu, Yeon-Koo Che, and Yosuke Yasuda 2011). I don't try to resolve this issue. Sections V.C-D consider design when level- k -incentive-compatibility is required and Section VI relaxes it.

C. General value densities with observable levels

In this section I consider general well-behaved value densities, assuming that traders' levels are observable or predictable, so that the mechanism can be exactly tailored to the combination of their levels.¹⁸ For ease of notation I also assume that each trader population has only one level.

As in MS's analysis, with ex post expected budget balance, the 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. For any mechanism (p, x) , let $f^k(v; p, x)$ and $F^k(v; p, x)$ be 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 $[0, 1]$, $f^1(v; p, x) \equiv 1$ 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 $g^2(c; p, x) \equiv g(\alpha_1^{-1}(c; p, x))$. Higher-level beliefs are defined analogously. I suppress dependence of $f^2(v; p, x)$ and $g^2(c; p, x)$ on (p, x) when 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 :

$$\begin{aligned}
 X_B^k(v) &= \int_0^1 x(v, \hat{c}) g^k(\hat{c}) d\hat{c}, & X_S^k(c) &= \int_0^1 x(\hat{v}, c) f^k(\hat{v}) d\hat{v}, \\
 (5.1) \quad P_B^k(v) &= \int_0^1 p(v, \hat{c}) g^k(\hat{c}) d\hat{c}, & P_S^k(c) &= \int_0^1 p(\hat{v}, c) f^k(\hat{v}) d\hat{v}, \\
 U_B^k(V, v) &= VP_B^k(v) - X_B^k(v), & U_S^k(C, c) &= X_S^k(c) - CP_S^k(c).
 \end{aligned}$$

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

$$(5.2) \quad U_B^k(V, V) \geq U_B^k(V, v) = VP_B^k(v) - X_B^k(v) \quad \text{and} \quad U_S^k(C, C) \geq U_S^k(C, c) = X_S^k(c) - CP_S^k(c).$$

¹⁸ Throughout the analysis I treat any differences in traders' levels as pure differences of opinion, as in Eliaz and Spiegler (2008): Traders neither believe others are better or worse informed nor draw inferences from others' actions or the chosen mechanism.

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

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

Theorems 2 and 3 extend MS's (their Theorems 1-2) characterization of mechanisms that maximize traders' expected total gains from trade in the set of equilibrium-incentive-compatible and interim individually rational mechanisms, to level- k models with observable traders' levels.

Theorem 2. *Assume that traders' levels are observable, i for the buyer and j for the seller. Then, for any mechanism that is incentive-compatible for traders of those levels,*

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

And if $p(\cdot, \cdot)$ is any function mapping $[0, 1] \times [0, 1]$ into $[0, 1]$, there exists a function $x(\cdot, \cdot)$ 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

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

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:

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

(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)), note that

$$(5.7) \quad \int_0^1 \int_0^1 (V - C) p(V, C)g(C)f(V)dCdV \\ = U_B^i(0,0) + \int_0^1 \int_0^V P_B^i(\hat{v})d\hat{v}f(V)dV + U_S^j(1,1) + \int_0^1 \int_C^1 P_S^j(\hat{c})d\hat{c}g(C)dC \\ = U_B^i(0,0) + U_S^j(1,1) + \int_0^1 [1 - F(V)] P_B^i(V)dV + \int_0^1 G(C) P_S^j(C)dC \\ = U_B^i(0,0) + U_S^j(1,1) + \int_0^1 \int_0^1 [G(C)f^j(V) + \{1 - F(V)\}g^i(C)] p(V, C)dCdV.$$

(5.4) implies (5.5) when the mechanism is interim individually rational for traders' levels. Given (5.3) and the monotonicity of $P_B^j(\cdot)$ and $P_S^k(\cdot)$, arguments like MS's (pp. 270-271) show that the analogue of MS's transfer function,

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

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

Before stating Theorem 3, consider whether MS's Corollary 1 generalizes when traders are level- k with observable levels. MS's Corollary 1 shows that if traders' values have positive probability densities over $[0,1]$, no equilibrium-incentive-compatible and equilibrium-interim individually rational trading mechanism can assure ex post efficiency with probability 1. If the buyer is level i and the seller is level j , and $p(V, C) \equiv 1$ iff $V \geq C$, the constraint (5.5) reduces to:

$$(5.9) \quad \begin{aligned} 0 &\leq \int_0^1 \int_0^1 \left(\left[V - \frac{\{1 - F(V)\}g^i(C)}{f(V)g(C)} \right] - \left[C + \frac{G(C)f^j(V)}{g(C)f(V)} \right] \right) p(V, C)g(C)f(V)dCdV \\ &= \int_0^1 \int_0^V [(V - C)g(C)f(V) - \{1 - F(V)\}g^i(C) - G(C)f^j(V)]dCdV \\ &= \int_0^1 \{F(V) - 1\}G^i(V)dV + \int_0^1 \{f(V) - f^j(V)\} \int_0^V G(C)dCdV. \end{aligned}$$

The first term on the right-hand side of (5.9) is analogous to MS's term, and with level- k beliefs $G^i(V)$; it is again always negative. The second term vanishes for correct beliefs and so has no counterpart in MS's analysis. To see that the second term can outweigh the first, consider LI beliefs: $f^j(V) \equiv 1$ and $g^i(C) \equiv 1$. Then, e.g., $F(\cdot)$ and $G(\cdot)$ with full supports, but with $F(\cdot)$ approximately uniform on $[b, 1]$ and $G(\cdot)$ with an approximate spike at b , make the right-hand side of (5.9) positive, a contradiction that shows MS's Corollary 1 does not generalize. But (5.9) shows that level- k incentive constraints still tend to limit the probability of ex post efficient trade.

Theorem 3 gives a concrete characterization of mechanisms that maximize traders' expected total gains from trade in the set of mechanisms that are incentive-compatible and interim individually rational for the traders' observed levels. Define, for $\beta \geq 0$,

$$(5.10) \quad \begin{aligned} \psi^{ij}(V, C; \beta) &= \left[V - \beta \frac{\{1 - F(V)\}g^i(C)}{f(V)g(C)} \right] - \left[C + \beta \frac{G(C)f^j(V)}{g(C)f(V)} \right] \\ &= (V - C) - \beta \left[\frac{\{1 - F(V)\}g^i(C)}{f(V)g(C)} + \frac{G(C)f^j(V)}{g(C)f(V)} \right], \text{ and} \\ p_\beta^{ij}(V, C) &= 1 \text{ if } \psi^{ij}(V, C; \beta) \geq 0, \text{ and } p_\beta^{ij}(V, C) = 0 \text{ if } \psi^{ij}(V, C; \beta) \leq 0. \end{aligned}$$

If feasible, $p_0^{ij}(V, C)$ would yield an ex post efficient allocation; but it may violate the incentive budget constraint (5.5). By contrast, $p_1^{ij}(V, C)$ maximizes the slack in (5.5); but it wastes expected gains from trade. The goal is an optimal compromise between these two extremes.

Theorem 3. *Assume that traders' levels are observable, i for the buyer and j for the seller. If there exists a level- k -incentive-compatible and interim individually rational 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' expected total gains from trade among all mechanisms that are incentive-compatible and interim individually rational for traders' levels. Further, 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). Consider choosing $p(\cdot, \cdot)$ to maximize traders' expected total gains from trade subject to $0 \leq p(\cdot, \cdot) \leq 1$, $U_B^i(0,0) = U_S^j(1,1) = 0$, and (5.5). (5.5) and (5.10) yield:

$$(5.11) \quad \begin{aligned} & \max_{\{0 \leq p(\cdot, \cdot) \leq 1\}} \int_0^1 \int_0^1 (V - C) p(V, C) g(C) f(V) dC dV \\ & \text{s.t. } 0 \leq \int_0^1 \int_0^1 \Psi^{ij}(V, C; 1) p(V, C) g(C) f(V) dC dV. \end{aligned}$$

If a solution $p(\cdot, \cdot)$ to problem (5.11) happens to yield a $P_B^i(\cdot)$ that is weakly increasing for all v and a $P_S^j(\cdot)$ that is weakly decreasing for all c , then by Theorem 2 that solution is associated with a mechanism that maximizes traders' expected total gains from trade among all level- k -incentive-compatible and interim individually rational mechanisms.

Problem (5.11) is like a consumer's budget problem, with the trade probabilities $p(V, C)$ like a continuum of goods and with "prices" $\left[\frac{\{1-F(V)\}g^i(C)}{f(V)g(C)} + \frac{G(C)f^j(V)}{g(C)f(V)} \right]$. Because the $p(V, C)$ enter the objective function and 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 (analogous to highest marginal-utility-to-price ratios).

Form the Lagrangean:

$$(5.12) \quad \begin{aligned} & \int_0^1 \int_0^1 (V - C) p(V, C) g(C) f(V) dC dV + \lambda \int_0^1 \int_0^1 \Psi^{ij}(V, C; 1) p(V, C) g(C) f(V) dC dV \\ & = \int_0^1 \int_0^1 \left(V - C + \lambda \Psi^{ij}(V, C; 1) \right) p(V, C) g(C) f(V) dC dV \\ & = (1 + \lambda) \int_0^1 \int_0^1 \left(\Psi^{ij}(V, C; 1) \frac{\lambda}{1 + \lambda} \right) p(V, C) g(C) f(V) dC dV. \end{aligned}$$

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:

$$(5.13) \quad (1 + \lambda)\Psi^{ij}\left(V, C; \frac{\lambda}{1+\lambda}\right) \leq 0 \text{ or equivalently}$$

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

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

$$(5.14) \quad (1 + \lambda)\Psi^{ij}\left(V, C; \frac{\lambda}{1+\lambda}\right) \geq 0 \text{ or equivalently}$$

$$(V - C) - \frac{\lambda}{1+\lambda} \left[\frac{\{1-F(V)\}g^i(C)}{f(V)g(C)} + \frac{G(C)f^j(V)}{g(C)f(V)} \right] \geq 0, \text{ when } p(V, C) = 1.$$

Given the continuity and monotonicity of $\Psi^{ij}(V, C; \beta)$, there are a unique λ and $p = p_\beta^{ij}(V, C)$, with $\beta = \frac{\lambda}{1+\lambda}$ (or $\lambda = \frac{\beta}{1-\beta}$), 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 conditions that (in my notation from (2.8)) $\Phi(V, 1)$ and $\Gamma(C, 1)$ are increasing on $[0, 1]$, which are 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 $g(C)$, then Theorem 3's condition would reduce to MS's condition. With level- k beliefs, Theorem 3's condition jointly restricts the beliefs and true densities in a qualitatively 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 maximize expected total gains from trade in the set of level- k -incentive-compatible and interim individually rational mechanisms, like MS's mechanisms that maximize expected gains from trade subject to the equilibrium incentive constraints, never require commitment to ex post perverse trade for any values. However, also as in MS's analysis, the transfer function (5.8) may require payment even from buyers who don't get the object.

Theorems 2 and 3 show that when level- k -incentive-compatibility and interim individually rationality are required and traders' levels are observable, MS's characterization of mechanisms that maximize expected gains from trade subject to the equilibrium incentive constraints for general well-behaved value densities generalizes, qualitatively, to the level- k model. The proofs show that the analysis goes through for any model that respects decision-theoretic rationality and makes generically unique predictions that satisfy monotonicity restrictions like MS's.

Comparing the level- k incentive budget constraint (5.5) with (2.5) (MS's (2)) and (5.14) with MS's (p. 274) condition (2.8) shows that design features that enhance a mechanism's performance in MS's analysis also enhance it in the level- k analysis, but with different weights.

The level- k analysis identifies another such feature, via (5.14). Unless a mechanism that maximizes expected total gains from trade subject to level- k incentive constraints happens to induce correct beliefs (as 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. Relative to a mechanism that maximizes traders' expected total gains from trade subject to equilibrium incentive constraints, TEPIB favors trade at (V, C) combinations for which traders' beliefs make the level- k "prices" $\left[\frac{\{1-F(V)\}g^i(C)}{f(V)g(C)} + \frac{G(C)f^j(V)}{g(C)f(V)} \right]$ lower than MS's equilibrium "prices" $\left[\frac{1-F(V)}{f(V)} + \frac{G(C)}{g(C)} \right]$. For $L2$ s and higher levels (excluding $L1$ s because their beliefs are exogenous), TEPIB also favors mechanisms that increase the advantages of the first two effects.

As in MS's analysis, mechanisms that maximize expected total gains from trade subject to level- k incentive constraints can be solved for in closed form only with uniform value densities, in which case they reduce to MS's direct mechanism for that case (Theorem 1). To assess how level- k thinking affects design, Figure 1 reports the trading regions for mechanisms that maximize expected total gains from trade subject to level- k and equilibrium incentive constraints, for a subset of the possible combinations of linear value densities.¹⁹ The equilibrium and $L1$ trading regions are generally similar. For densities that make $L1$ s' beliefs pessimistic (optimistic), the $L1$ trading regions are usually supersets (subsets) of equilibrium trading regions.

D. General value densities with unobservable levels

In this subsection I assume that traders' levels are unobservable, but that the designer knows that the population level distributions for both buyers and sellers include, with positive prior probabilities, both $L1$ and one or more higher levels. I continue to require level- k -incentive-

¹⁹ 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.

compatibility and interim individual rationality for all levels with positive probability, and I assume that traders can draw no inferences from the mechanism (see footnote 18).

I begin with a useful lemma, which first appeared as part of the proof of Theorem 1 in early versions of this paper (see also Gorelkina 2018 and de Clippel et al. 2019).

Lemma. *A mechanism is level- k incentive-compatible and interim individually rational for all levels if and only if it is both $L1$ - and equilibrium-incentive-compatible and $L1$ - and equilibrium-interim individually rational.*

Proof. The “only if” part is trivial. The “if” part follows because if a mechanism is $L1$ -incentive-compatible and interim individually rational, $L2$ s best respond to $L1$ s’ predicted truthful reports, so $L2$ s’ incentive-compatibility and interim individual rationality constraints (5.2)-5.3) and (5.5)-(5.6) coincide with equilibrium traders’ incentive-compatibility and interim individual rationality constraints (2.2)-(2.3) and (2.5)-(2.6) (MS’s (2)-(4)). By induction, the coincidence extends to higher levels ad infinitum. Q.E.D.

A random posted-price mechanism (Hagerty and Rogerson 1987; Čopič and Ponsati 2008, 2016) is a distribution over prices π and a probability density $\mu(\cdot)$ such that trade occurs at price π with probability $\mu(\pi)$ if $v \geq \pi \geq c$, with no trade or transfer otherwise. A deterministic posted-price mechanism is one for which the density $\mu(\cdot)$ is concentrated on a single price.

Under conditions on traders’ value densities that are generically satisfied, Theorem 4 characterizes mechanisms that maximize traders’ expected total gains from trade among all level- k -incentive-compatible and interim individually rational mechanisms.²⁰

Theorem 4. *Assume that traders’ levels are unobservable and the population level distributions for buyers and sellers are both known to include, with positive prior probabilities, $L1$ and one or more higher levels. Assume that traders’ value distributions $F(\cdot)$ and $G(\cdot)$ are such that $f(\cdot)$, $g(\cdot) \neq 1$ almost everywhere and there is no x for which $F(x) = x$ and $G(x) = x$. Then, 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 all level- k -incentive-compatible and interim individually rational mechanisms must be equivalent to a deterministic posted-price mechanism with $U_B^i(0,0) = U_S^j(1,1) = 0$ for all levels*

²⁰ Larry Samuelson and Rene Saran noted errors in previous proofs of Theorem 4 and made important suggestions that led to this proof. A similar result holds, with a similar proof, if there are multiple levels with positive probability with the highest higher than $L1$; but I give the result for $L1$ and some higher levels for concreteness and ease of notation.

i and j in the trader populations, $p(v, c) = 0$ or 1 almost everywhere, and an optimal posted price π uniquely characterized by the first-order condition:

$$(5.15) \quad \frac{f(\pi)}{g(\pi)} = \frac{\int_{\pi}^1 (V-\pi)f(V)dV}{\int_0^{\pi} (\pi-C)g(C)dc} = \frac{E(V-\pi|V \geq \pi)}{E(\pi-C|C \leq \pi)}.$$

Remark. Recall that Theorem 1 shows that with uniform value densities, even with heterogeneous, unobserved levels, MS's mechanism that maximizes traders' expected total gains from trade subject to equilibrium incentive constraints also maximizes traders' expected total gains from trade subject to level- k incentive constraints. Theorem 4's conditions on $F(\cdot)$ and $G(\cdot)$ rule out both uniform value densities and non-generic densities that mimic uniform densities on the supports of any discrete set of posted prices (as in Wolfgang Leininger, Peter B. Linhart, and Roy Radner 1989, pp. 76-78). In those cases, different levels' incentive-compatibility and individual rationality constraints coincide; but in all other cases they differ. Theorem 4's conditions are plainly stronger than needed, but are generically satisfied.

Proof. Given that with positive probabilities the buyer population includes levels 1 and i and the seller population includes levels 1 and j , (5.6) must hold for those levels by Theorem 2. The Lemma shows that if a mechanism is $L1$ -incentive-compatible, it is Lk -incentive-compatible for all $k > 1$ if and only if it is also equilibrium-incentive-compatible. That and (5.1) imply that:

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

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

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

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

Standard arguments show that (5.16)-(5.19) are almost everywhere differentiable in V or C . As $f(\cdot)$ and $g(\cdot)$ are continuous, differentiability fails only where $p(v, c)$ is discontinuous, so $p(v, c)$ is almost everywhere continuous. $p(v, c)$ is plainly weakly monotonic in the obvious directions. Given (2.1), (2.6), (5.1), (5.6), and levels 1 and i buyers' and levels 1 and j sellers' utilities from trading the object, their expected utilities given V and C , including transfers, are:

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

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

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

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

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 of (5.22) and (5.23) with respect to C must each be equal whenever they exist. Thus, after cancellations we must have almost everywhere (with subscripts denoting partial differentiation):

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

Because $p(v, c)$ is weakly monotonic and $f(\cdot), g(\cdot) \neq 1$ almost everywhere, (5.24) implies that $p(v, c)$ is locally independent of v and c almost everywhere. Thus $p(v, c)$ must be a step function—in v for any given c , or vice versa—with at most a countable number of steps. If there is more than one step, buyers and sellers must each choose the local posted price associated with their true values, and those buyers and sellers on an edge between steps must be indifferent. But those indifference conditions are monotonic in higher levels' beliefs, and given that there is no x for which $F(x) = x$ and $G(x) = x$, such indifferences cannot hold for all of them. It follows that any level- k -incentive-compatible mechanism has at most one step, and thus that it is a (possibly random) posted-price mechanism: “possibly” because traders make their reports after the posted price is drawn. Such mechanisms make truthful reporting a dominant strategy for any $k > 0$.

Finally, we can take $p(v, c) = 0$ or 1 almost everywhere without loss of generality because the $p(v, c)$ enter the objective function and constraints linearly. An optimal deterministic posted-price mechanism would choose the price π to solve:

$$(5.25) \quad \max_{\{0 \leq \pi \leq 1\}} \int_{\pi}^1 \int_0^{\pi} (V - C) p(V, C) g(C) f(V) dC dV.$$

The second-order condition of problem (5.25) is satisfied globally, so a random posted price is never optimal. The uniquely optimal deterministic posted price is characterized by (5.15). Q.E.D.

Theorem 4's optimal posted-price mechanism is independent of traders' value and level distributions and so can be implemented without knowledge of them. That it makes truthful reporting a dominant strategy for all levels, so level- k and equilibrium incentive constraints coincide, is a plausible rationale for the dominant-strategy implementation often assumed in robust mechanism design (Hagerty and Rogerson 1987; Čopič and Ponsatí 2008, 2016).

The optimal posted price balances expected gains and losses from π 's effect on the values for which trade occurs. π is determined by true expected-surplus tradeoffs, independent of traders' beliefs, so in this case there are no TEPIB benefits. The resulting static mechanism comes close to satisfying Wilson's (1987) desideratum, in that its *rules* are distribution-free. However, the optimal price is determined, via (5.15), by conditional means that depend on the value densities.

Čopič and Ponsatí (2016) describe a dynamic, continuous-time double auction in which the auctioneer reveals bids to traders only once they are compatible, which can implement any posted-price mechanism without knowledge of the densities, satisfying Wilson’s desideratum.

A posted price rules out the sensitive dependence on reported values of the mechanisms that maximize expected total gains from trade for traders whose levels are observable. Theorem 1 allows a rough estimate of the cost in gains from trade of giving up such sensitive dependence: Theorem 1’s optimal mechanism with uniform value densities yields probability of trade $9/32 \approx 28\%$ and expected gains from trade $9/64 \approx 0.14$. By contrast, the optimal posted price with uniform value densities is $1/2$, which yields probability of trade $1/4 = 25\%$ and expected surplus $1/8 = 0.125$, a seemingly modest cost for the robustness of a posted-price mechanism.

As already noted, Theorem 4 is close to Kneeland’s (2018) and De Clippel et al.’s (2019) results for the case of heterogeneous, unobservable levels. Mine are specific to MS’s bilateral trading environment. But Theorem 4’s result would continue to hold for any sufficiently well-behaved model (including De Clippel et al.’s 2019 wider set of LO anchors) that respects decision-theoretic rationality and makes generically unique predictions. It would also make De Clippel et al.’s equal treatment of levels a conclusion rather than an assumption.

VI. RELAXING LEVEL- k -INCENTIVE-COMPATIBILITY

This section relaxes Section V’s requirement of level- k -incentive-compatibility while retaining interim individual rationality, still allowing only direct mechanisms. Here one can still define a general class of feasible direct mechanisms, with payoff-relevant outcomes $p(v, c)$ and $x(v, c)$. However, such mechanisms’ effects can no longer be tractably captured via incentive constraints like (5.2) and (5.6), and must be modeled directly via level- k traders’ responses. I briefly consider the cases where traders’ levels are observable and unobservable in turn.

A. Observable levels

I assume uniform value densities for simplicity. 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 level- k traders continue to anchor beliefs on an LO that is uniformly 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 that restricted menu instead of $[0, 1]$, and such anchoring can make reserve prices useful.²¹

For example, $L1$ traders in a double auction without reserve prices believe they face bids or asks uniformly distributed on $[0, 1]$, yielding outcomes that do not maximize expected total gains from trade subject to $L1$ incentive constraints. In a double auction with reserve prices for buyer's bids of $\frac{3}{4}$ and seller's asks of $\frac{1}{4}$, if $L1$ traders anchor on the restricted menu, they bid or ask as if facing asks or bids uniformly distributed on $[\frac{1}{4}, 1]$ or $[0, \frac{3}{4}]$ respectively, or equivalently (given the ranges of their optimal bids or asks) on $[\frac{1}{4}, \frac{3}{4}]$ for both: exactly the ranges of serious bids or asks in CS's linear double-auction equilibrium (Section I). A double auction with those reserve prices therefore rectifies $L1$ traders' beliefs and is outcome-equivalent to MS's mechanism that maximizes expected gains from trade for equilibrium traders. The probability of trade is $9/32 \approx 28\%$ and the expected total gains from trade are $9/64 \approx 0.14$ (Section II.C), far higher than $L1$ s' probability of trade $1/8 = 12.5\%$ and expected gains $1/24 \approx 0.04$ in the double auction without reserve prices (Section IV.A). Pushing the reserve prices beyond $\frac{3}{4}$ and $\frac{1}{4}$ further reduces the value gap needed for trade, which is a benefit, other things equal; but it also precludes some bids or asks needed for trade. The cost of precluding those bids or asks turns out to exceed the benefits, and computations show that reserve prices of $\frac{3}{4}$ and $\frac{1}{4}$ are in fact optimal.

For $L2$ s, with uniform value densities, a double auction without reserve prices already improves upon MS's mechanism that maximizes expected gains from trade for equilibrium traders, or a mechanism that maximizes expected gains in the set of $L2$ -incentive-compatible and interim individually rational mechanisms (Sections II.C, IV.B, V.A). Feasible reserve prices (restricted to $[0, 1]$) bring $L2$ s' beliefs closer to equilibrium beliefs, eliminating some of the unaggressiveness 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 probability of trade $25/72 \approx 35\%$ and expected total gains from trade $11/72 \approx 0.15$, higher than the equilibrium probability of trade $9/32 \approx 28\%$ and expected gains from trade $9/64 \approx 0.14$.

²¹ I know of no evidence for 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 shows that reserve prices are not useful in this setting if equilibrium is assumed. Crawford et al. (2009) showed that in first-price auctions, if level- k bidders anchor on a restricted menu of bids, reserve prices can be useful even though they would be useless with equilibrium bidders.

B. Unobservable levels

Turning to the case of unobservable levels, Section V.C-D's results allow a rough estimate of the potential benefits of allowing non-level- k -incentive-compatible direct mechanisms. As already noted, exact implementation is not always the most natural approach to level- k design. Suppose, for example, that the population is known to include multiple levels with the frequency of one of them high. Then a mechanism that maximizes expected gains from trade in the set of level- k -incentive-compatible and interim individually rational mechanisms for only that level, or possibly a non-level- k -incentive-compatible mechanism (Section VI.A), will yield more surplus than a mechanism that maximizes expected total gains from trade in the set of level- k -incentive-compatible and interim individually rational mechanisms for the actual distribution of levels.

Specifically, with uniform value densities, the mechanism that maximizes expected gains from trade subject only to $L1$ incentive constraints yields probability of trade $9/32 \approx 28\%$ and expected surplus $9/64 \approx 0.14$ (Section II.C); and its performance will approach these values continuously as the frequency of $L1$ s approaches one. By contrast, the mechanism that maximizes expected gains from trade subject to the incentive constraints for all levels is a posted-price mechanism with optimal price $1/2$, probability of trade $1/4 = 25\%$, and expected total surplus $1/8 = 0.125$, independent of the population level frequencies.

If $L2$'s frequency is high enough, the double auction without reserve prices yields probability of trade $25/72 \approx 35\%$ and expected surplus $11/72 \approx 0.15$, an even larger increase (Section IV.B).

More generally, relaxing the restriction to level- k -incentive-compatible mechanisms can yield optimal mechanisms that differ qualitatively as well as quantitatively from those that are optimal for equilibrium traders, with substantial increase in performance.

VII. CONCLUSION

This paper has revisited Myerson and Satterthwaite's (1983; "MS") analysis of mechanism design for bilateral trading with independent private values, replacing their equilibrium assumption with the assumption that traders follow a structural nonequilibrium model based on level- k thinking and restricting attention to direct mechanisms. The level- k model makes specific predictions that allow an analysis of mechanism design with power comparable to MS's equilibrium analysis. The results clarify the role of their equilibrium assumption in several ways.

The anchoring of level- k beliefs on $L0$ creates level- k menu effects (Crawford et al. 2009) that make the revelation principle fail. It thus matters whether level- k -incentive-compatibility is

truly required. If it is required, much of MS's analysis is qualitatively robust to relaxing equilibrium in favor of a level- k or another kind of structural nonequilibrium model.

With uniform value densities, MS's closed-form solution for the direct mechanism that maximizes equilibrium traders' expected total gains from trade subject to incentive constraints remains valid for any population of level- k traders—though only if that mechanism is implemented in its incentive-compatible form, not as the double auction. With general well-behaved value densities, if traders' levels are observable or predictable, MS's characterization of mechanisms that maximize total expected gains from trade subject to equilibrium incentive constraints extends qualitatively to level- k models. Design features that enhance mechanism performance in MS's analysis then do so for level- k traders, with different weights. The level- k analysis adds a novel enhancing feature, tacit exploitation of predictably incorrect beliefs.

To put these conclusions another way, the equilibrium assumption bundles decision-theoretic rationality with homogeneity and statistical correctness of traders' beliefs, and the revelation principle sidesteps the issue of uniqueness. The level- k analysis shows that if traders' levels are observable or predictable, one can largely dispense with homogeneity, statistical correctness, and uniqueness, which are behaviorally the least plausible of MS's assumptions and which at first sight appear to play essential roles in their analysis.

If traders' levels are unobservable, the mechanism must generally screen traders' levels along with their values. Then, except in special cases such as that of uniform value densities, level- k -incentive-compatibility compels the use of a (possibly random) posted-price mechanism, and the mechanism that maximizes traders' expected total gains from trade must be a particular deterministic posted-price mechanism, which I characterize. Such a mechanism can be implemented dynamically, following Čopić and Ponsatí (2016), without knowledge of the details of the environment, in a way that fully satisfies Wilson's (1987) desideratum.

A final section briefly considers the implications of failures of the revelation principle's implication that it does not matter whether incentive-compatibility is required. If *non*-level- k -incentive-compatible direct mechanisms are truly feasible, violating level- k -incentive-compatibility can yield mechanisms that differ qualitatively as well as quantitatively from those that maximize expected total gains from trade when equilibrium is assumed, with large 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

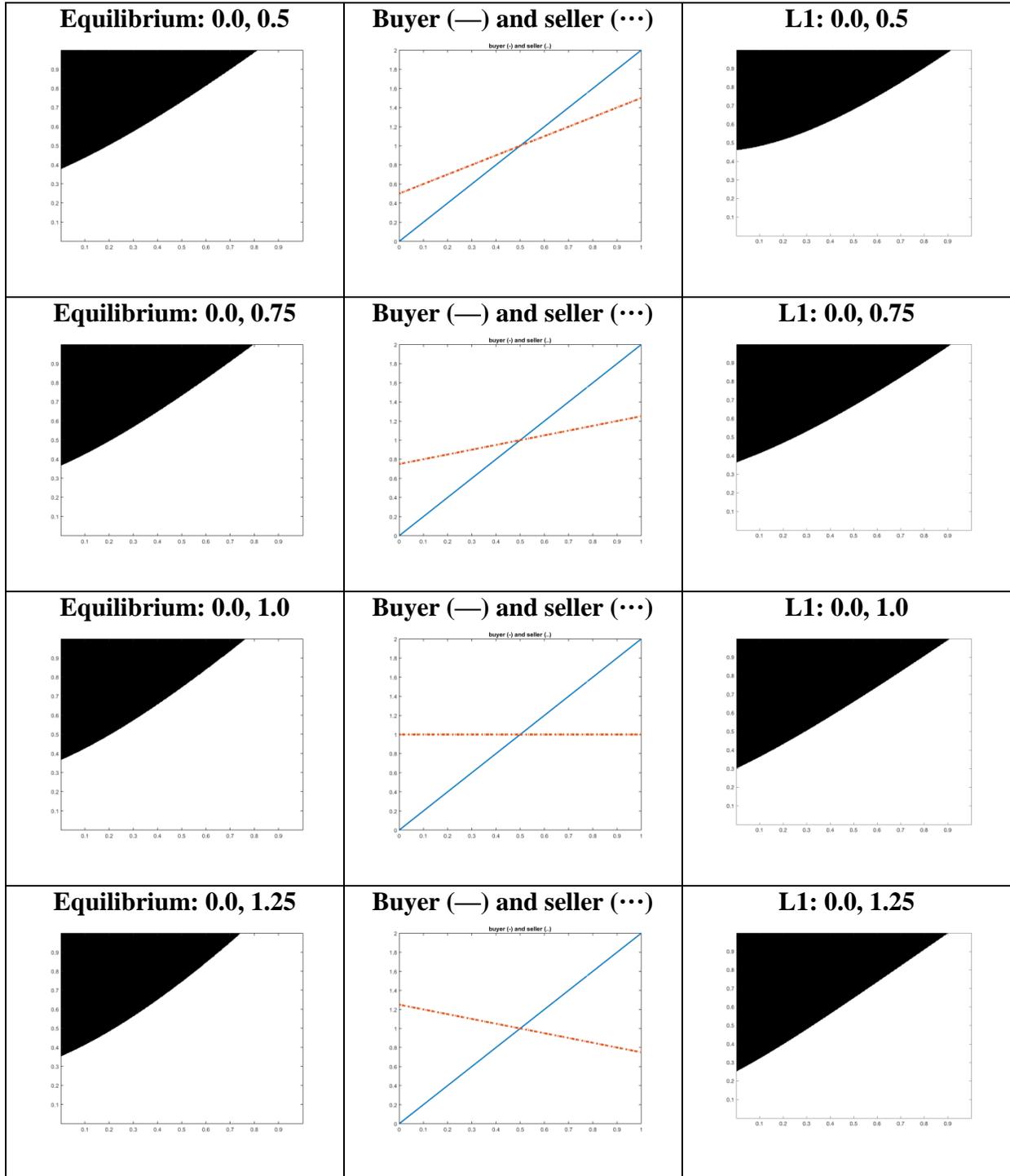
- Abdulkadiroglu, Atila, Yeon-Koo Che, and Yosuke Yasuda (2011): "Resolving Conflicting Preferences in School Choice: The 'Boston Mechanism' Reconsidered," *American Economic Review*, 101, 1–14.
- Abdulkadiroglu, Atila, and Tayfun Sönmez (2003): "School Choice: A Mechanism Design Approach," *American Economic Review*, 93, 729-747.
- Agranov, Marina, Elizabeth Potamites, Andrew Schotter, and Chloe Tergiman (2012): "Beliefs and Endogenous Cognitive Levels: An Experimental Study," *Games and Economic Behavior*, 75, 449-463.
- Aumann, Robert J., and Adam Brandenburger (1995): "Epistemic Conditions for Nash Equilibrium," *Econometrica*, 63, 1161-1180.
- Bartling, Björn, and Nick Netzer (2016): "An Externality-robust Auction: Theory and Experimental Evidence," *Games and Economic Behavior*, 97, 185-204.
- Bergemann, Dirk, and Stephen Morris (2009): "Robust Implementation in Direct Mechanisms," *Review of Economic Studies*, 76, 1175-1204.
- 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.
- Bierbrauer, Felix, and Nick Netzer (2016): "Mechanism Design and Intentions," *Journal of Economic Theory*, 163, 557-603.
- Börger, Tilman, and Jiangtao Li (forthcoming): "Strategically Simple Mechanisms," *Econometrica*, in press.
- 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 (2016): “Optimal Robust Bilateral Trade: Risk Neutrality,” *Journal of Economic Theory*, 163, 276–287.
- 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 (2016): “Mechanism Design with Bounded Depth of Reasoning and Small Modeling Mistakes”, manuscript, August 2016; <https://drive.google.com/file/d/0B8Sv4TBdx30Jc0JRRUpkb3EzVTQ/edit?usp=sharing>.
- de Clippel, Geoffroy, Rene Saran, and Roberto Serrano (2019): “Level- k Mechanism Design,” *Review of Economics Studies*, 86, 1207-1227.
- Eliasz, Kfir, and Ran Spiegler (2006): “Contracting with Diversely Naïve Agents,” *Review of Economic Studies*, 73, 689–714.
- Eliasz, Kfir, and Ran Spiegler (2007): “A Mechanism-Design Approach to Speculative Trade,” *Econometrica*, 75, 875–884.
- Eliasz, 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.
- Gill, David, and Victoria Prowse (2016): “Cognitive Ability, Character Skills, and Learning to Play Equilibrium: A Level- k Analysis,” *Journal of Political Economy*, 124, 1619-1676.
- 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 (2018): “The Expected Externality Mechanism in a Level- k Environment,” *International Journal of Game Theory*, 47, 103–131.
- 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 (2018): “Mechanism Design with Level- k Types: Theory and an Application to Bilateral Trade,” manuscript;
http://www.tkneeland.com/uploads/9/5/4/8/95483354/levelk_mechanismdesign_24.10.2018.pdf
- Leininger, Wolfgang, Peter B. Linhart, and Roy Radner (1989), “Equilibria of the Sealed-Bid Mechanism for Bargaining with Incomplete Information,” *Journal of Economic Theory*, 48, 63-106.
- 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.
- Pathak, Parag (2017): “What Really Matters in Designing School Choice Mechanisms,” in Bo Honore, Ariel Pakes, Monika Piazzesi, and Larry Samuelson, eds., *Advances in Economics and Econometrics*, 11th World Congress of the Econometric Society, Cambridge U. Press.
- 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.
- Saran, Rene (2016): “Bounded Depths of Rationality and Implementation with Complete Information,” *Journal of Economic Theory*, 165, 517–564.
- 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.
- Williams, Steven R. (1987): “Efficient Performance in Two Agent Bargaining,” *Journal of Economic Theory*, 41, 154–172.
- 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 (2016): “Mechanism Design with Maxmin Agents: Theory and an Application to Bilateral Trade,” *Theoretical Economics*, 11, 971–1004.

Figure 1. Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of $L1$ -incentive-compatible mechanisms²²



²² 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 $f(1) = 2-x$, and the seller's density $g(C)$ satisfies $g(0) = y$ and $g(1) = 2-y$.

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

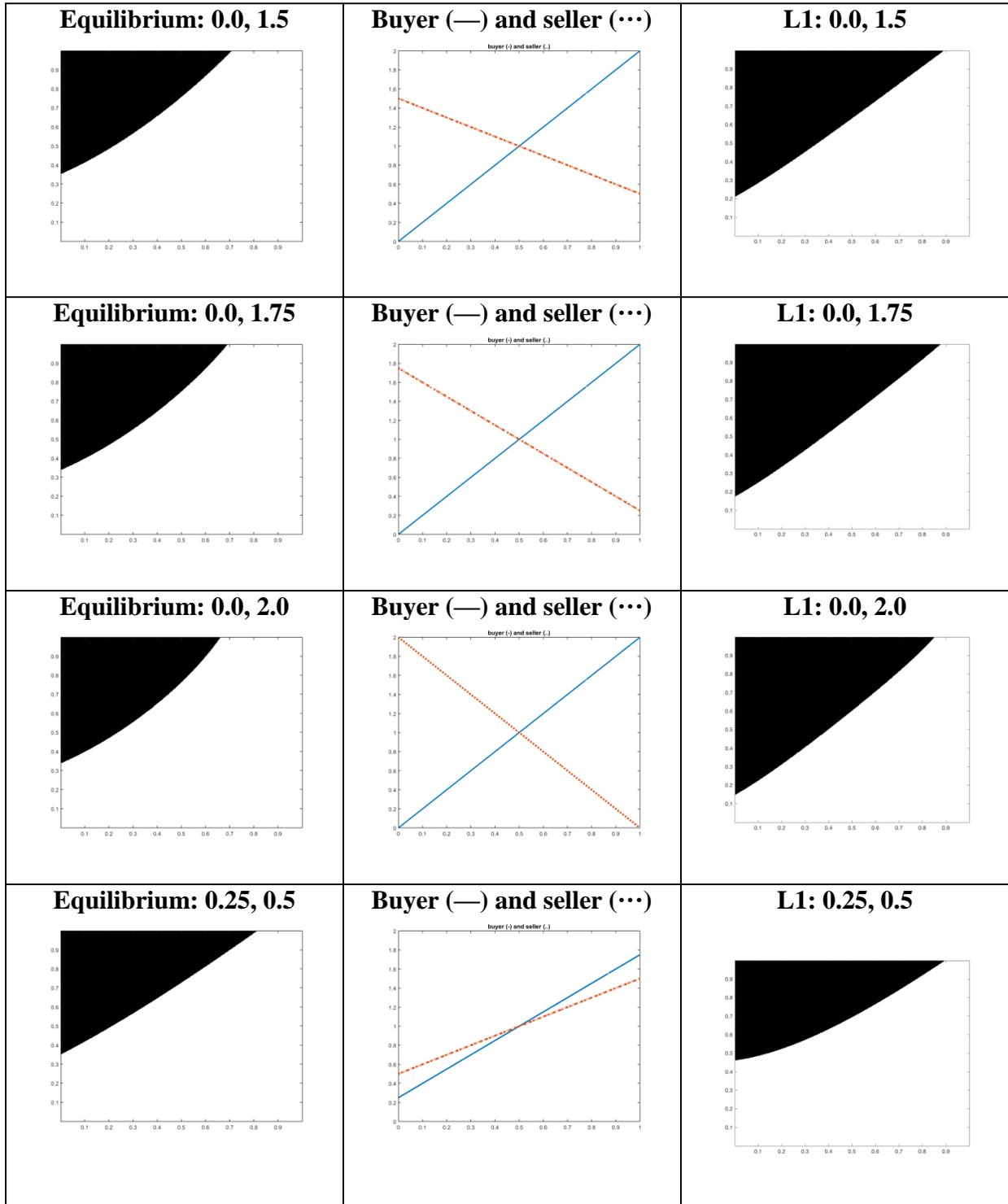


Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of $L1$ -incentive-compatible mechanisms

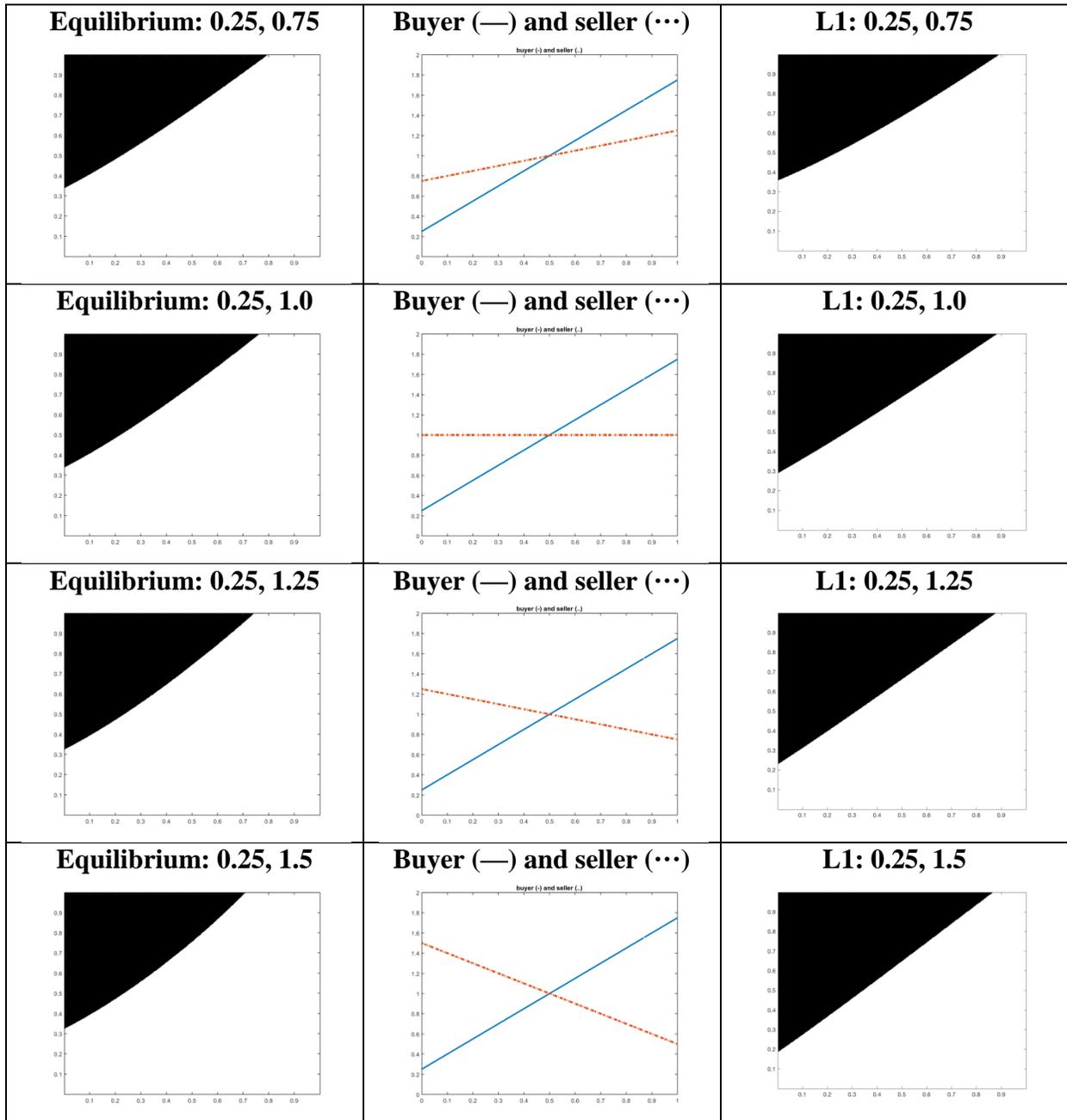


Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of $L1$ -incentive-compatible mechanisms

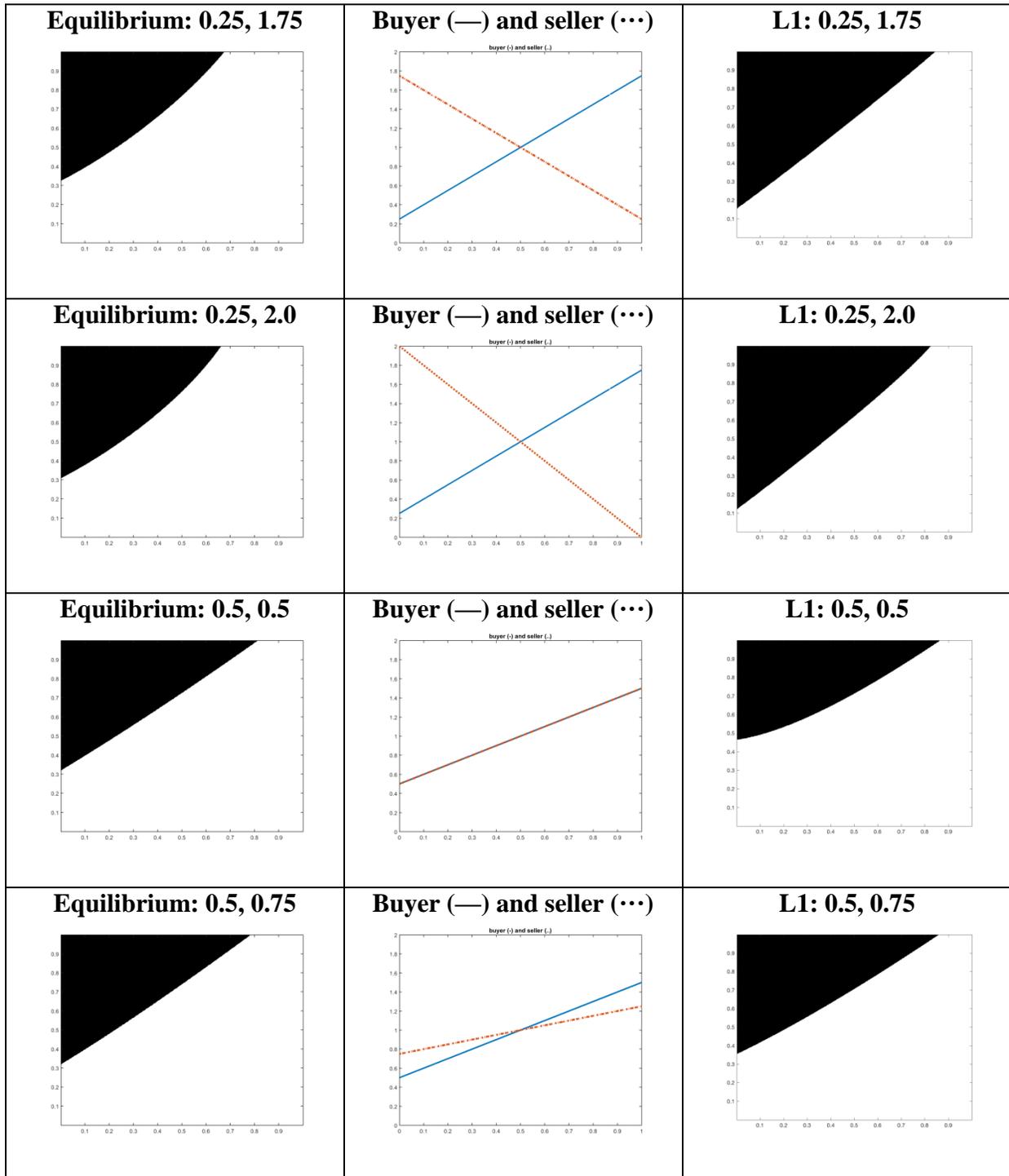


Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of $L1$ -incentive-compatible mechanisms

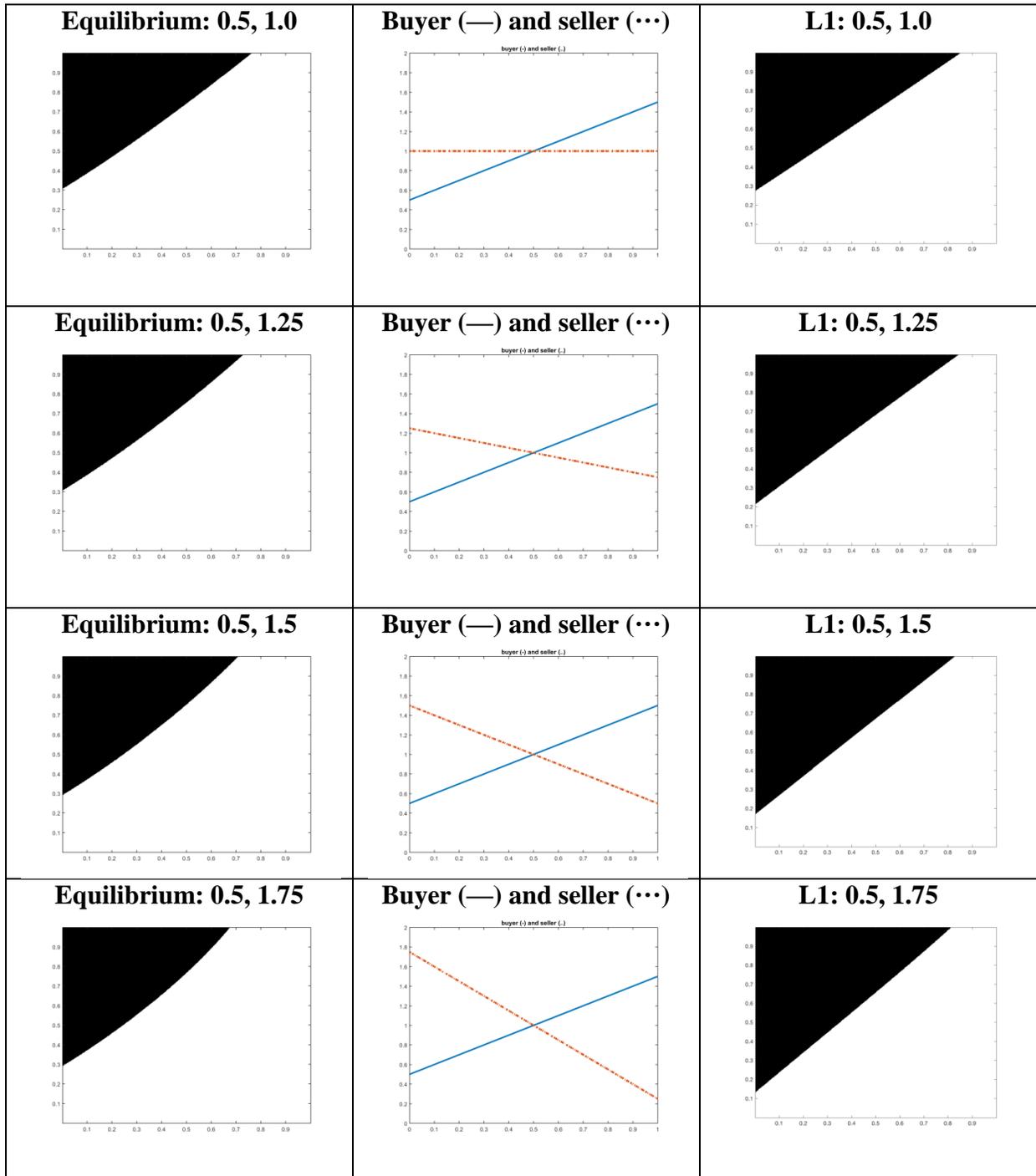


Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of $L1$ -incentive-compatible mechanisms

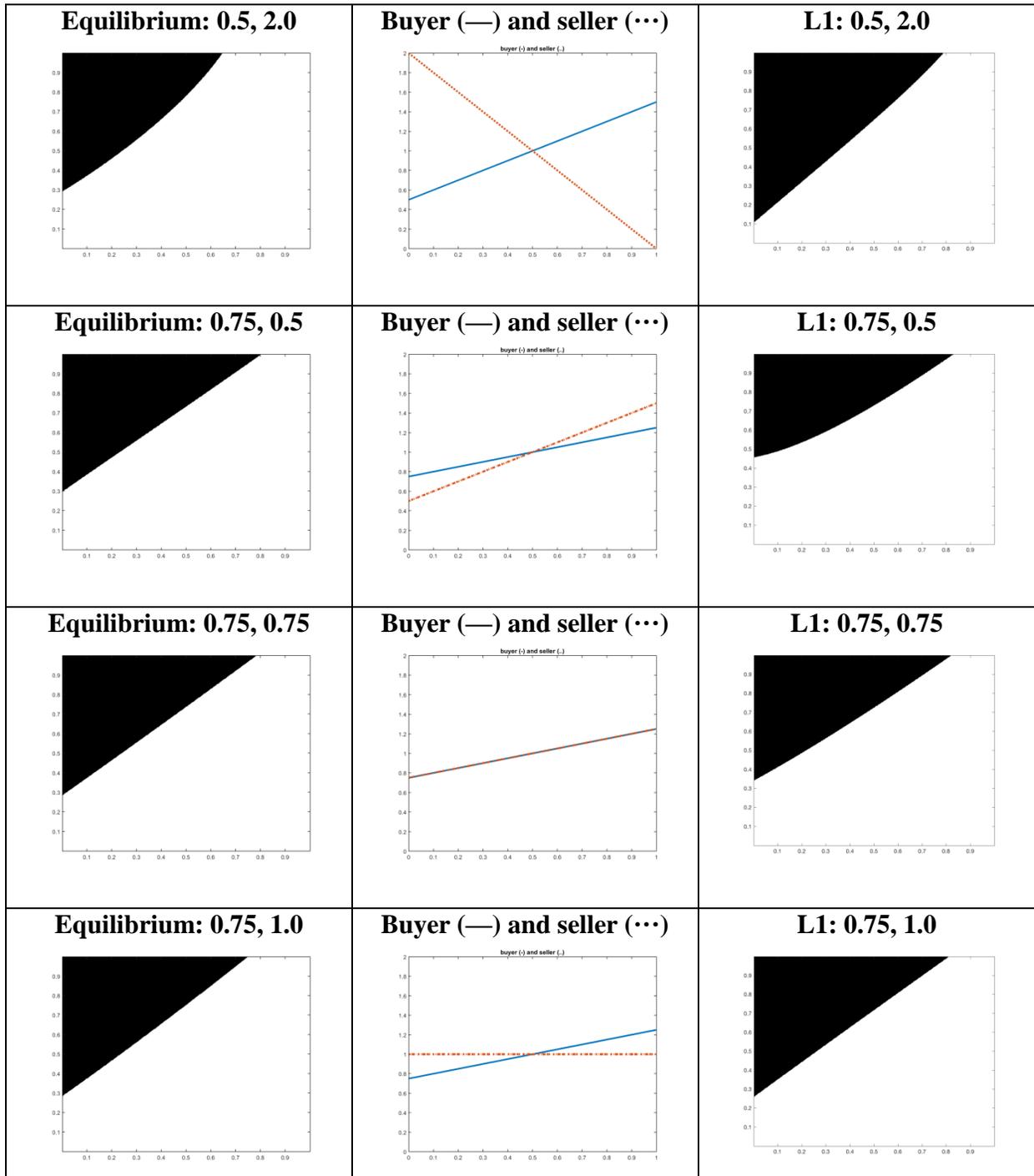


Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of $L1$ -incentive-compatible mechanisms

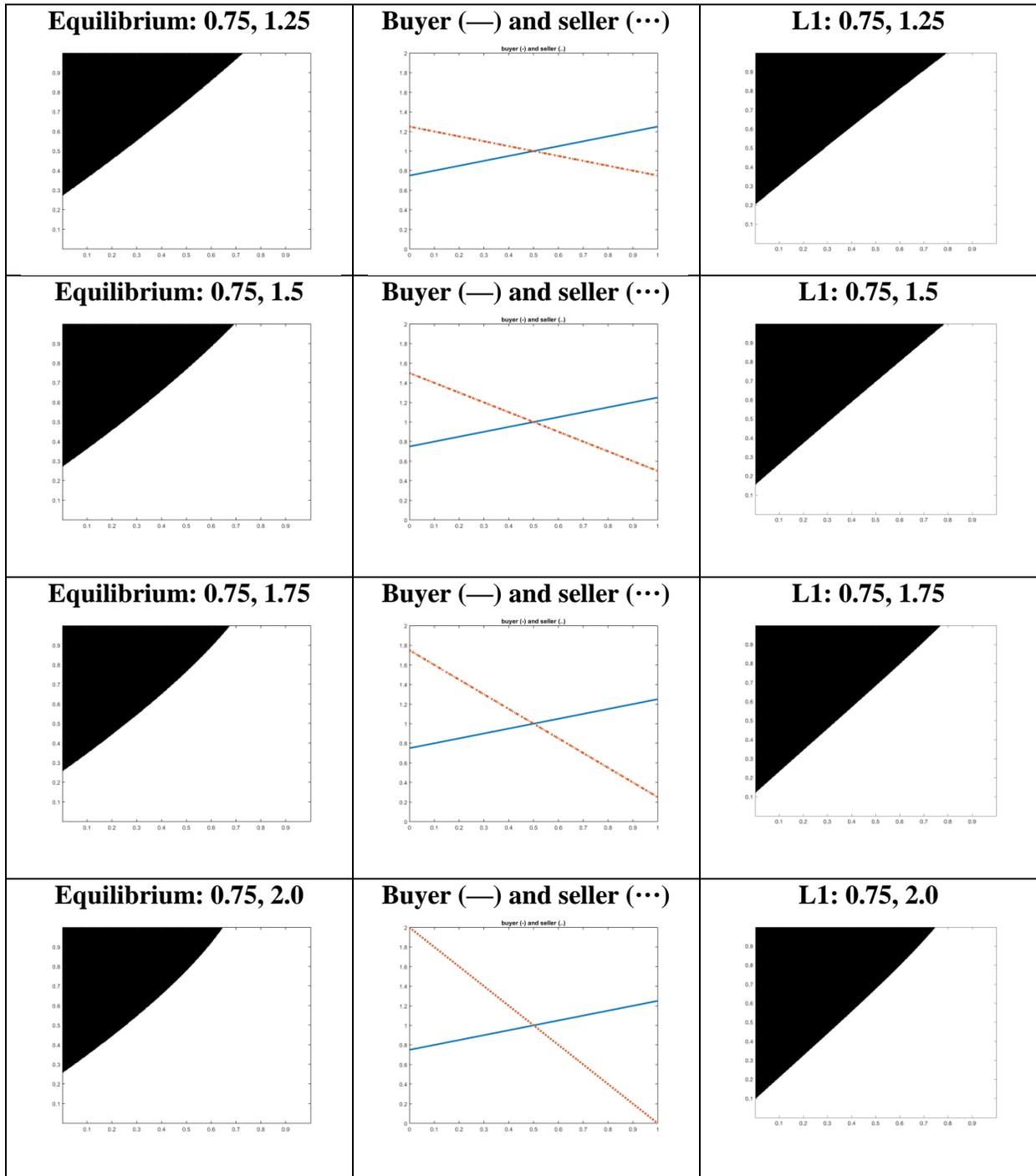


Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of $L1$ -incentive-compatible mechanisms

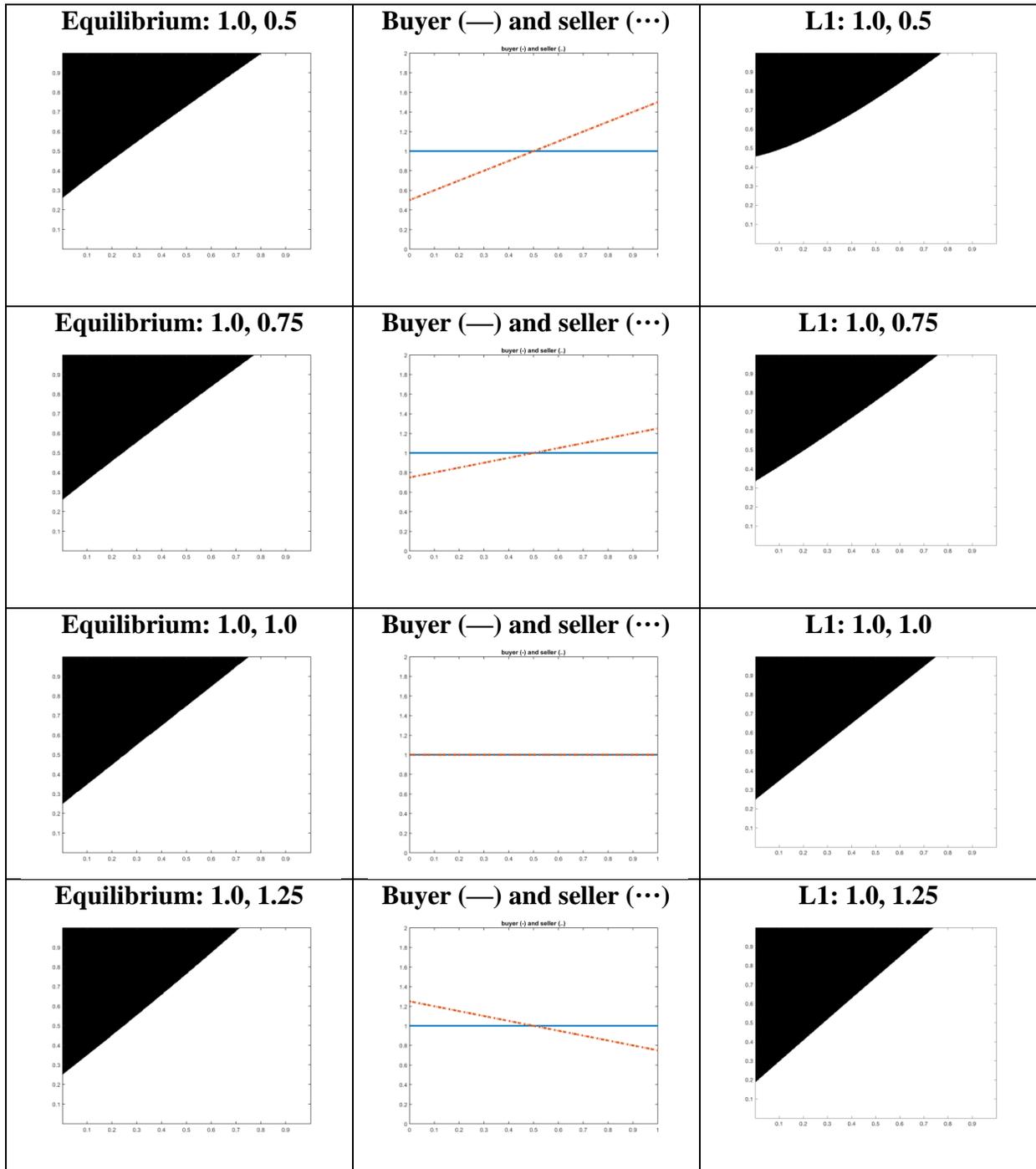


Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of $L1$ -incentive-compatible mechanisms

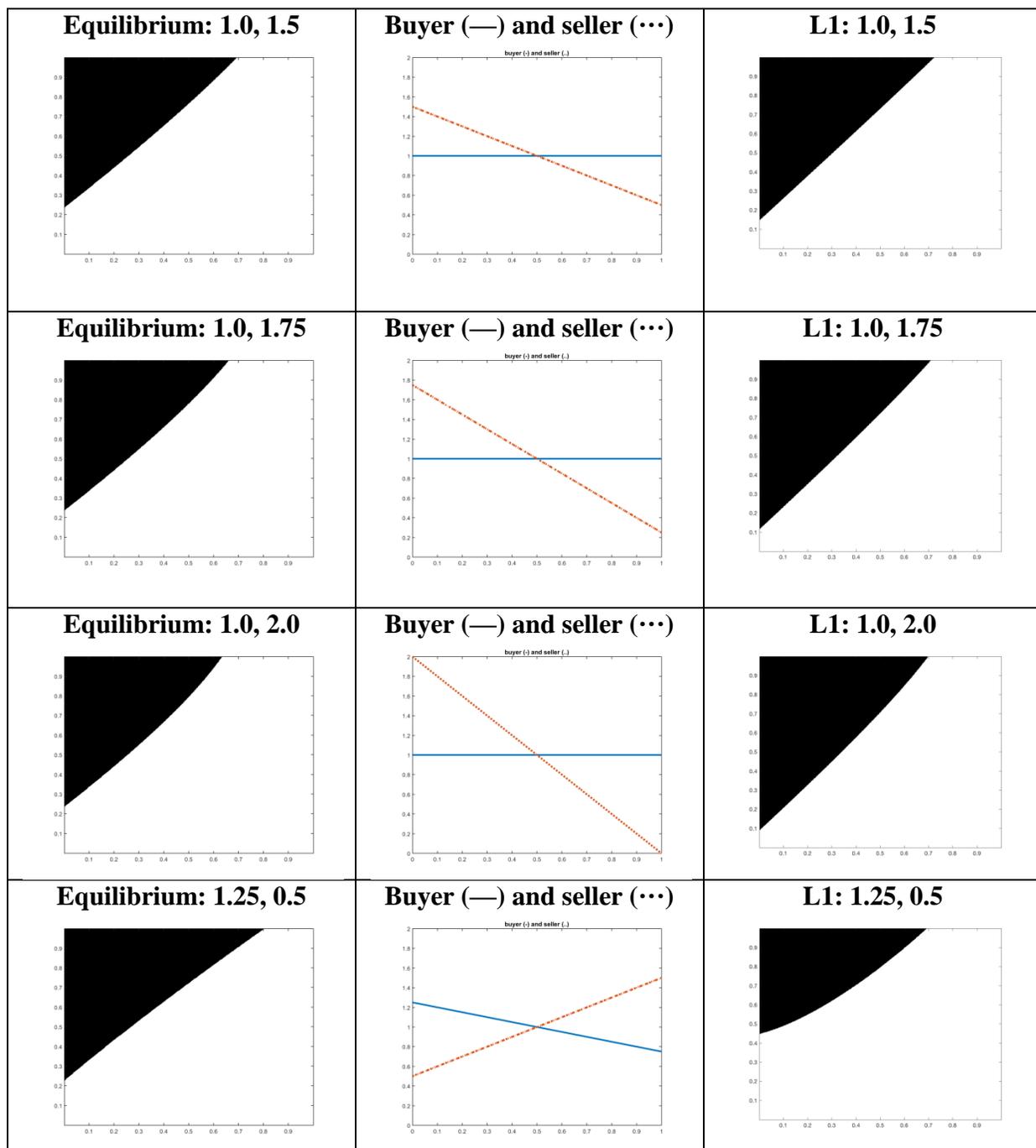


Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of $L1$ -incentive-compatible mechanisms

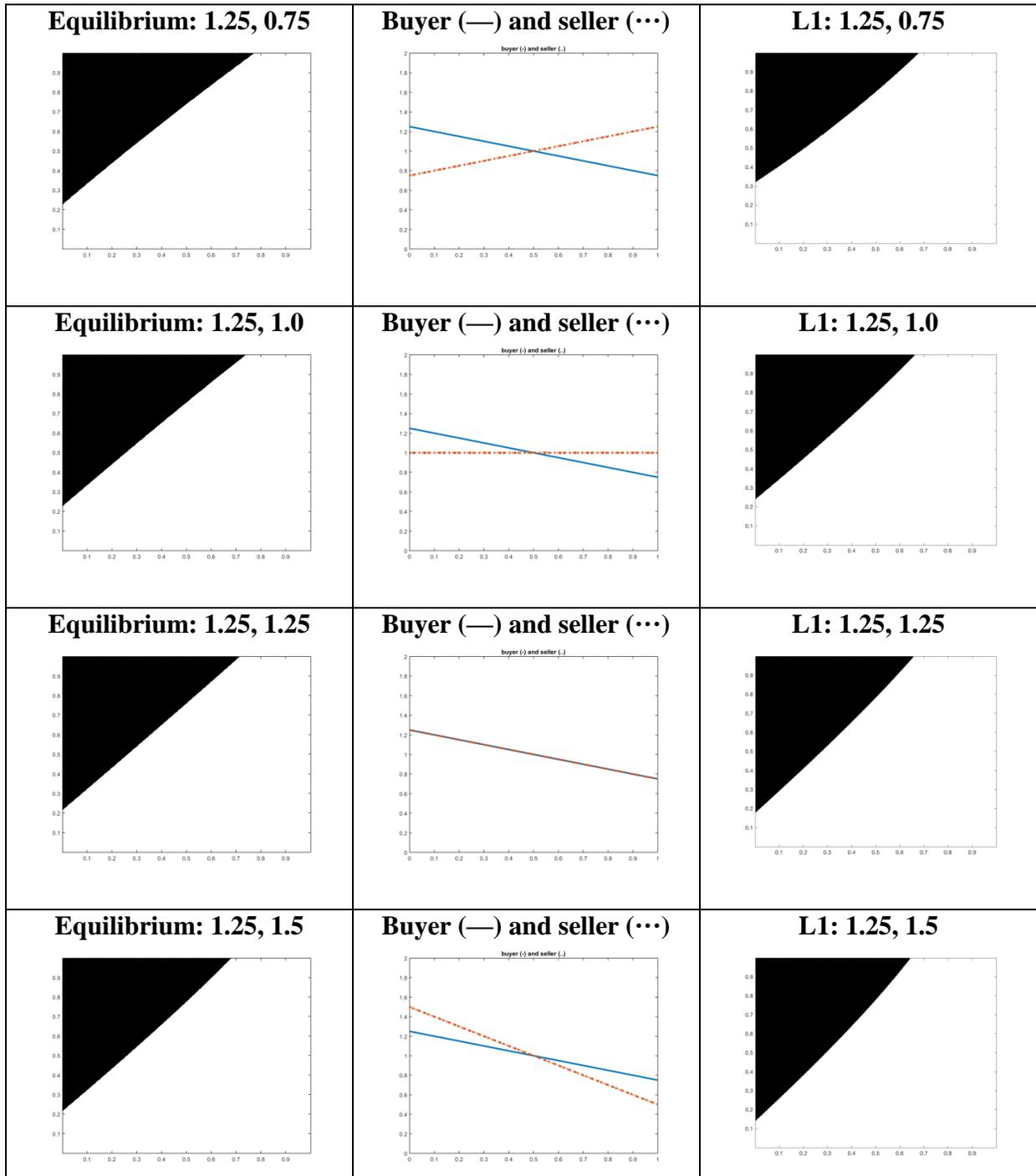


Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of $L1$ -incentive-compatible mechanisms

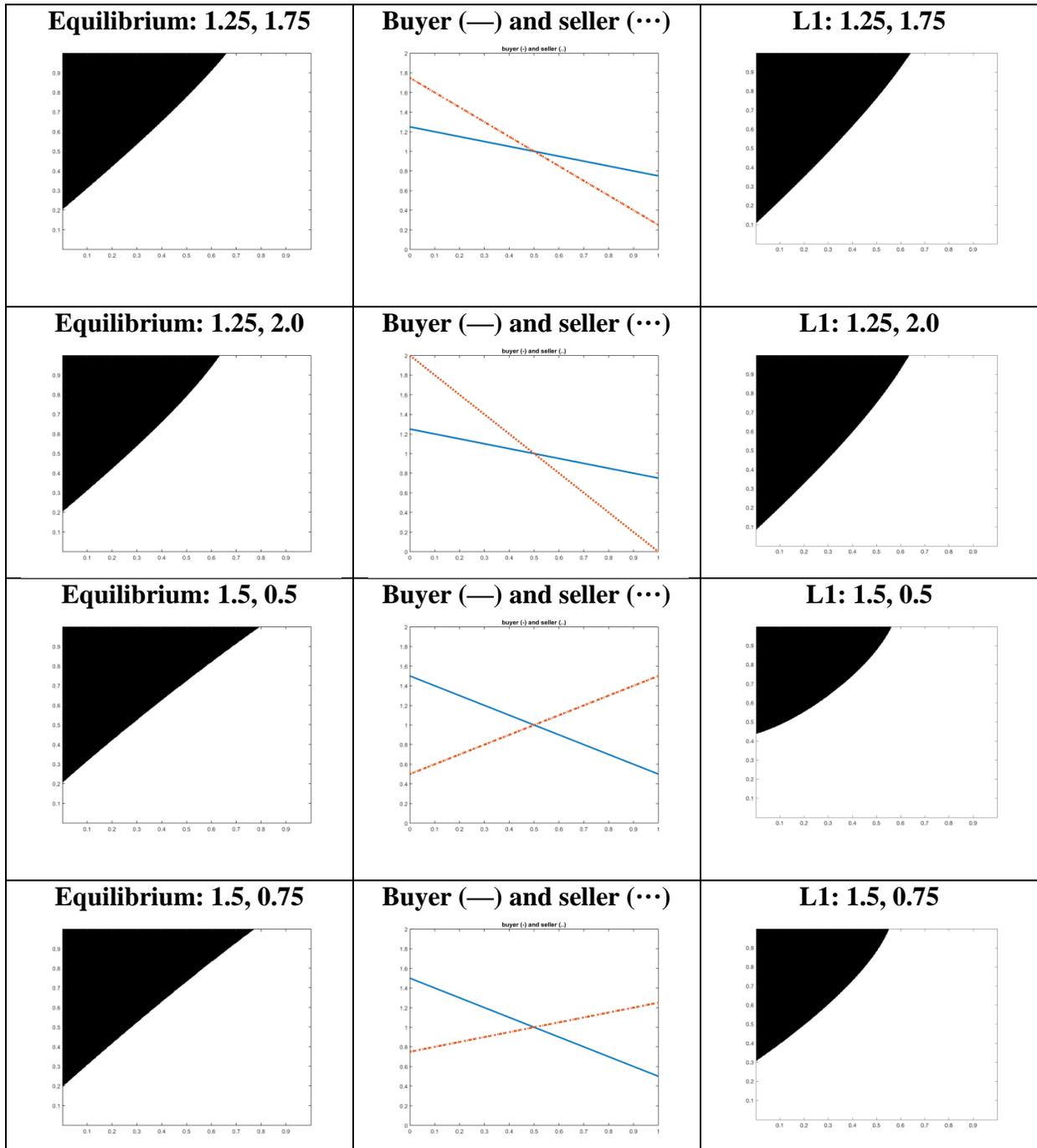


Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of $L1$ -incentive-compatible mechanisms

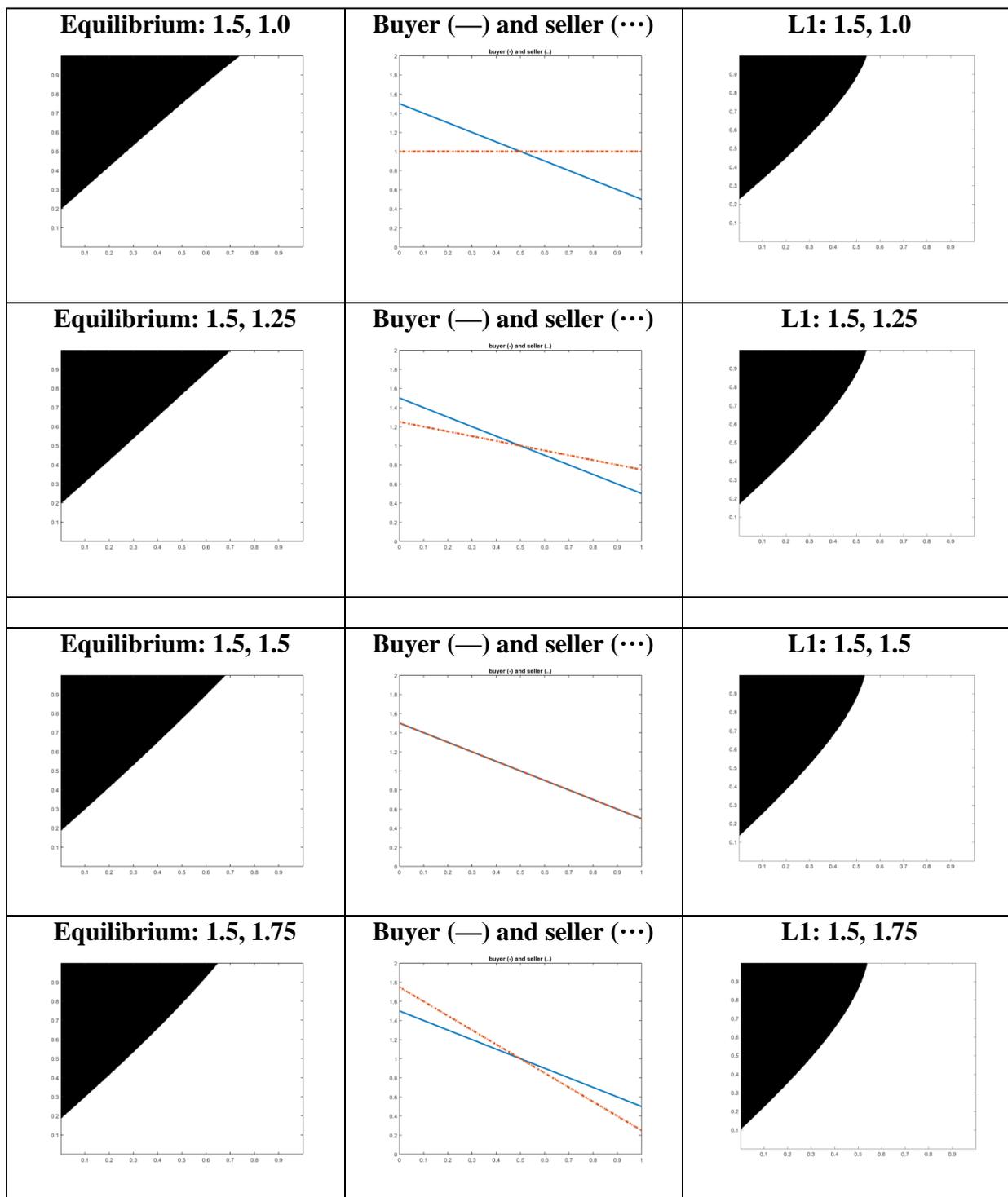
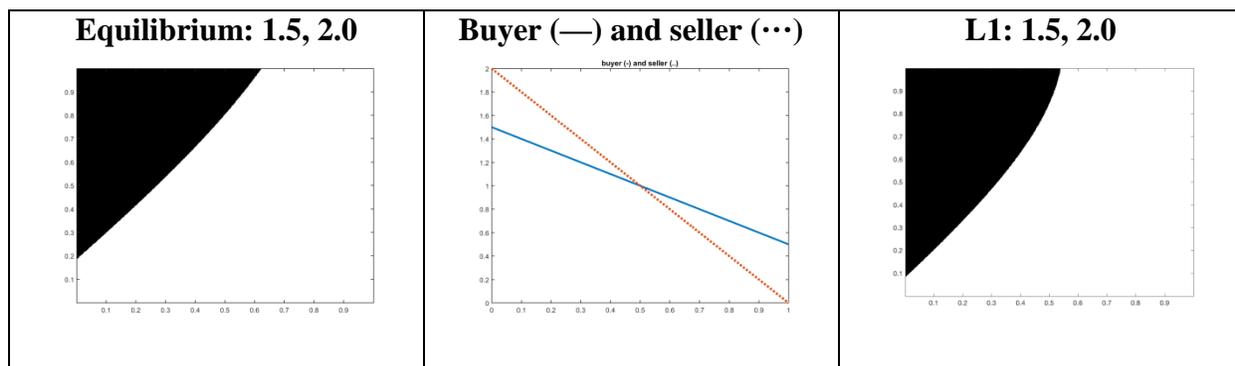


Figure 1 (cont.). Trading regions (in black) for equilibrium-incentive-efficient mechanisms and mechanisms that are efficient in the set of $L1$ -incentive-compatible mechanisms



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

The code assumes a homogeneous population of *LIs* 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 i th point. For instance, `fbar = 2` means $f(0) = 0.25$ and `gbar = 3` means $g(0) = 0.5$.

Solution algorithm

For *LIs*, 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 *LIs*) for that λ . It then iterates these operations, increasing λ by increments of 0.01, until it finds the λ that makes the value on the right-hand side of (5.5) smallest; and checks that that value does not change sign more than. 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 pair of value densities. Figure 1 is based on all possible discretized combinations of linear value densities.

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

Using the code to implement the algorithm

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 *LI* 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.

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 length.

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=0.01; % beginning value for the discretised value range of lambda
alfafin=1;    % ending value for the discretised value range of lambda
alfaincr=0.01; % increment value for the discretised value range of lambda

v=beg:incr:fin; % generation of discretised values of buyer
c=beg:incr:fin; % generation of discretised values of seller

fbar=charbeg:charincr:charfin; % generation of discretised characterising
                               % values of buyer's value distribution
gbar=charbeg:charincr:charfin; % generation of discretised 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))-((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))+((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)...
+(phis-gammas)*pi*fdist*gdist*norm ;
% integration

sumvekk (fbars,gbars,alfas)=sumvekk(fbars,gbars,alfas)...
+(phiss-gammass)*pieq*fdist*gdist*norm ;
% integration-for eqm counterpart

```

```

        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);
        namell = strcat(name1,name2,'L1');
        nameeqm = strcat(name1,name2,'eqm');

        l1= blackandwhite(fbars,gbars,pi);
        eqm=blackandwhites(fbars,gbars,pis);

        saveas(l1,namell,'png');
        saveas(eqm,nameeqm,'png');
        yeni (fbars,gbars,fbar,gbar);

    end
end

```

```

toc;

```

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
```

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= (v(vs)/gdist)-((1-fcum)/(fdist*gdist));
```

```
z= (c(cs)/fdist)+(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= (1-fcum)/(fdist);
```

```
    z= 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(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'} );

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
```