Efficient Mechanisms for Level-*k* Bilateral Trading[*]

Vincent P. Crawford

Department of Economics and All Souls College, University of Oxford, UK

Department of Economics, University of California, San Diego, La Jolla, CA, USA

First version 18 April 2015; this revision 27 January 2021

Abstract: This paper revisits Myerson and Satterthwaite's (1983) classic analysis of mechanism design for bilateral trading, replacing equilibrium with a level-*k* model of strategic thinking and focusing on direct mechanisms. The revelation principle fails for level-*k* models, so restricting attention to direct mechanisms and imposing incentive-compatibility are not without loss of generality. If, however, only direct, level-*k*-incentive-compatible mechanisms are feasible and traders' levels are observable, Myerson and Satterthwaite's characterization of mechanisms that maximize traders' total surplus subject to incentive constraints generalizes qualitatively to level-*k* models. If only direct, level-*k*-incentive-compatible mechanisms are feasible but traders' levels are not observable, generically a particular posted-price mechanism maximizes traders' total expected surplus subject to incentive constraints. If direct, non-level-*k*-incentive-compatible mechanisms are feasible and traders best respond to them, total expected surplus-maximizing mechanisms may take completely different forms.

Keywords: mechanism design, bilateral trading, level-*k* thinking, behavioral game theory

JEL codes: C70, D02

[*] Preliminary versions of this paper were presented as an Economic Journal Lecture at a Royal Economic Society Conference and a John von Neumann Distinguished Lecture at Brown University's 250th Anniversary Celebration. This version corrects significant errors in previous versions of Theorems 2-4. I owe thanks to Johannes Abeler, Douglas Bernheim, Colin Camerer, Yan Chen, Miguel Costa-Gomes, Jacques Crémer, Avinash Dixit, Kfir Eliaz, Tore Ellingsen, Xavier Gabaix, Olga Gorelkina, Nagore Iriberri, Paul Klemperer, Margaret Meyer, Paul Milgrom, Roger Myerson, Philip Neary, Zvika Neeman, Nick Netzer, Robert Östling, Ady Pauzner, Phil Reny, Adam Sanjurjo, Roberto Serrano, Joel Sobel, Alex Teytelboym, several anonymous referees, and especially Tilman Börgers, Larry Samuelson, and Rene Saran. Rustu Duran provided 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. Declarations of interest: none.

1. Introduction

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 people's initial responses to games. I also focus on direct mechanisms, those that elicit reports from traders that are conformable to estimates of their values. Otherwise I maintain standard assumptions about behavior and design.[1]

Equilibrium-based analyses of design have enjoyed tremendous success; and both theory and experiments support the assumption that players in a game who have had 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 be justified logically via epistemic arguments (Aumann and Brandenburger, 1995). However, in experiments that study initial responses to games, subjects' thinking seldom follows the fixed-point or iterated-dominance logic that equilibrium normally requires without learning.[2] Instead their thinking often favors level-*k* decision rules, which anchor subjects' 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 levels of thinking is normally concentrated on *L1*, *L2*, and perhaps *L3* (Costa-Gomes and Crawford, 2006; Crawford, Costa-Gomes, and Iriberri, 2013, Sections 3 and 5).

For $k > 0$, *Lk* is decision-theoretically rational, with an accurate model of the game. It departs from equilibrium only in basing its beliefs on an oversimplified model of others. *Lk*'s decisions

[1] As in MS's and almost all other analyses of design, I assume that traders' responses follow the behavioral model noiselessly.

[2] 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 using their own ways of thinking. 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 make equilibrium unique and/or discoverable via iterated dominance. However, experiments suggest that neither feature ensures equilibrium initial responses in the kinds of games used in most analyses of implementation (Katok, Sefton, and Yavas, 2002; Chen and Ledyard, 2008).

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]

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 specific predictions, 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 perform worse in practice than mechanisms that are more robust. Finally, a level-*k* analysis can reduce the sensitivity of theoretically optimal mechanisms to distributional and knowledge assumptions that observed institutions seldom respond to (Wilson, 1987).

The revelation principle fails for level-*k* models, due to "level-*k* menu effects" (Crawford, Kugler, Neeman, and Pauzner, 2009; "CKNP"). As a result, neither restricting attention to direct mechanisms nor imposing level-*k*-incentive-compatibility is without loss of generality, as they are in MS's equilibrium-based analysis. Even so, I restrict attention to direct mechanisms, which brings the analysis closer to mechanisms used in practice and makes it more concrete and arguably more informative.[5] I assume throughout that the population frequency of *L0* is zero, as most evidence suggests. I use "equilibrium" and "level-*k*" as needed to distinguish concepts that depend on traders' beliefs, such as incentive-compatibility and interim individual rationality.

The characterization of mechanisms that maximize level-*k* traders' total expected surplus depends on two things: whether only level-*k*-incentive-compatible and interim individually rational mechanisms are truly feasible and whether traders' levels can be observed.[6]

[3] In Camerer, Ho, and Chong's (2004) cognitive hierarchy model, *Lk* best responds to an estimated mixture of lower levels, including *L0*, with the level distribution constrained to be Poisson. *L2* and higher levels may not respect *k*-rationalizability. A cognitive hierarchy version of my analysis would be feasible, with some loss of clarity.

[4] Until recently the alternatives to assuming equilibrium were limited to quantal response equilibrium and rationalizability or *k*-rationalizability. To my knowledge, quantal response equilibrium has not been applied to design, perhaps because its predictions must be solved for numerically and are sensitive to its error structure. Rationalizability and *k*-rationalizability have been applied to design (Section 7); but the level-*k* model's precision allows an analysis that yields additional insight.

[5] Ollár and Penta (2017, 2019) give similar arguments for restricting attention to direct mechanisms. But de Clippel, Saran, and Serrano (2019) and Kneeland (2018) obtain useful results in nonequilibrium analyses that allow indirect mechanisms.

[6] Experimental evidence suggests that subjects' levels are correlated with observable variables, if not perfectly predictable (e.g. Agranov, Potamites, Schotter, and Tergiman, 2012; Alaoui and Penta, 2016; Alaoui, Janezic, and Penta, 2020). Even imperfect correlations sometimes influence applications (e.g. Pathak, 2017, Section 2.4.2). Here, however, I follow MS and the subsequent design literature in treating observability or predictability as an all-or-nothing distinction.

My main results are for cases where only level-*k*-incentive-compatible and level-*k*-interim individually rational mechanisms are feasible. Then, if traders' value densities are uniform, MS's equilibrium-based result that the incentive-compatible mechanism that mimics the linear equilibrium in Chatterjee and Samuelson's (1983; "CS") symmetric double auction also maximizes traders' total expected surplus subject to incentive constraints, generalizes exactly to level-*k* models with any distribution of levels, observable or not. Comparing that mechanism with level-*k* outcomes in the double auction reveals a failure of the revelation principle for level-*k* models: Unlike in MS's analysis, where the linear double-auction equilibrium is outcome-equivalent to their mechanism's truthful equilibrium, the total-surplus-maximizing level-*k*-incentive-compatible mechanism yields different outcomes than the double auction.

When only level-*k*-incentive-compatible and -interim individually rational mechanisms are feasible, if traders have general well-behaved value densities and their levels are observable—making their beliefs and best responses predictable so that the mechanism can be tailored to their combination of levels—then MS's characterization of mechanisms that maximize traders' total expected surplus subject to incentive constraints generalizes qualitatively to level-*k* models, with a novel feature, tacit exploitation of predictably incorrect beliefs. MS's result that no incentive-compatible mechanism can assure ex post efficient trade does not fully generalize in this case.

Still assuming that only level-*k*-incentive-compatible and interim individually rational mechanisms are feasible and that traders have general well-behaved value densities, if traders' levels are not observable, so that an incentive-compatible mechanism must screen levels and values simultaneously, then no fully responsive direct mechanism can screen them perfectly. Generically, screening both levels and values to maximize traders' total expected surplus subject to level-*k* incentive constraints requires a posted-price mechanism (Hagerty and Rogerson, 1987), whose limited responsiveness makes truthful revelation of values a weakly dominant strategy and therefore a best response for all levels. The optimal posted price is sensitive to traders' value densities, but it can be implemented dynamically without knowing them as in Čopič and Ponsatí (2008), in a way that satisfies Wilson's (1987) desiderata. Although a posted price is a coarser instrument than the mechanisms that maximize traders' total expected surplus when their levels are observable, the surplus cost of its robustness is modest, at least if the value densities are close to uniform. MS's ex-post-inefficiency result generalizes in this case.

Finally, if direct, non-level-*k*-incentive-compatible mechanisms are feasible and traders best respond to them, total expected-surplus-maximizing mechanisms may take quite different forms.

These results can be thought of as follows. At first glance, MS's analysis appears to depend on the full strength of their assumption that traders will play the desired equilibrium in any game that the designer's choice of mechanism creates. MS's use of equilibrium bundles four distinct behavioral assumptions: decision-theoretic rationality, homogeneity of strategic thinking, and predictability and coordination/correctness of beliefs. Experimental evidence suggests that homogeneity of strategic thinking and coordination/correctness of beliefs are behaviorally more questionable than the other two, especially in games of incomplete information like those CS and MS studied (Crawford and Iriberri, 2007; Brocas, Carrillo, Camerer, and Wang, 2014).

The level-*k* model unbundles those four behavioral assumptions, retaining decision-theoretic rationality while relaxing homogeneity of strategic thinking and coordination/correctness of beliefs in a structured way. The level-*k* model also links the predictability of traders' beliefs to the observability of their levels, relaxing predictability when levels are not observable, again in a structured way.[7] The relaxation traces the need for robust implementation to the unpredictability of strategic thinking, a plausible rationale for robustness. The analysis in that case brings total expected-surplus-maximizing mechanisms closer to the mechanisms used in applications, where the unpredictability of people's thinking often seems to exert a major influence on design.

The structure of the level-*k* model allows an analysis of cases where only direct, level-*k*-incentive-compatible mechanisms are feasible with most of the power and precision of MS's equilibrium-based analysis, retaining equilibrium's assumptions of decision-theoretic rationality and predictability or partial predictability of traders' beliefs, while dispensing with its strong assumptions of homogeneity of strategic thinking and coordination/correctness of beliefs.

Further, level-*k* analyses of non-level-*k*-incentive-compatible mechanisms promise to identify potential roles for indirect and/or non-incentive compatible mechanisms, which are irrelevant in an equilibrium analysis by revelation-principle fiat. Preliminary analyses suggest

[7] As the proofs suggest, many of my results would go through for other nonequilibrium models that respect rationality and make generically unique conditional predictions. Some, however, depend essentially on the level-*k* model's iterated-best-response structure. I conduct the analysis for level-*k* models because they are well supported by evidence, and for comprehensibility.

that nonequilibrium design will need to go well beyond implementing the best outcomes attainable in equilibrium under weaker behavioral assumptions.[8]

Section 2 reviews CS's equilibrium analysis of bilateral trading via double auction, the starting point for MS's analysis, while introducing the assumptions and notation regarding traders' preferences and information. Section 2 goes on to review MS's analysis of equilibrium design. Section 3 defines a level-*k* model for the incomplete-information games created by direct trading mechanisms. Section 4 extends CS's equilibrium analysis of the double auction to Section 3's level-*k* model. Section 5 considers level-*k* design when level-*k*-incentive-compatibility and interim individual rationality are required. Section 6 considers relaxing level-*k* incentive constraints, allowing direct mechanisms that create incentives to lie but assuming that traders best respond to them. Section 7 discusses related literature. Section 8 is the conclusion.

2. Equilibrium bilateral trading via double auction and equilibrium mechanism design

Following CS and MS, I consider bilateral trading between a potential seller and buyer of an indivisible object. 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" for both). 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 with no important loss of generality, I take their supports to be identical and normalize them to [0, 1].

2.1. Equilibrium bilateral trading via double auction

CS study a double auction, in which traders make simultaneous money offers. If the buyer's offer b (for "bid") exceeds the seller's offer a ("ask"), they exchange the object for a price that is a weighted average of a and b. CS allowed any weights from 0 to 1, but as in MS's analysis I focus on the symmetric case with weights ½. Then, if $b \geq a$, the buyer acquires the object at price $(a + b)/2$, the seller's utility is $(a + b)/2 - C$, 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 zero.

[8] CKNP illustrate this point concretely for auction design, showing that revenue-equivalence fails for level-*k* bidders, and that even though a second-price auction yields the equilibrium-optimal revenue level for level-*k* as well as equilibrium bidders, when the population is made up mostly of *L1* bidders a non-*L1*-incentive-compatible first-price auction yields more revenue.

CS show that this game has many Bayesian equilibria. In the leading case where traders' value densities $f(V)$ and $g(C)$ are uniform, CS derived a closed-form solution for a linear equilibrium, which plays an important role in MS's analysis. Denote the buyer's bidding strategy in that equilibrium $b_*(V)$ and the seller's asking strategy in that equilibrium $a_*(C)$. Then, with normalization of the supports of $f(V)$ and $g(C)$ to [0, 1], $b_*(V) = 2V/3 + 1/12$ 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. Thus traders' bids are shaded, so trade occurs if and only if $2V/3 + 1/12 \geq 2C/3 + 1/4$, or $V \geq C + 1/4$, at price $(V + C)/3 + 1/6$. In this and CS's other equilibria, the probability of an ex post inefficient outcome is positive.

2.2. Equilibrium mechanism design

MS ask whether the ex post inefficiency of CS's equilibria is an avoidable flaw of the double auction or rather a general property of any trading mechanism when traders' values are private. Efficient trading requires that traders' responses to the mechanism reveal their private values, at least implicitly. Conceivably, a different mechanism could accomplish that without inefficiency.

MS begin (pp. 267-268) with the revelation principle, the observation that any given equilibrium of the game created by a feasible mechanism can be viewed as the truthful equilibrium of some direct-revelation mechanism. Assuming equilibrium and the ability to select among equilibria, it follows that there is no loss of generality in restricting attention to direct mechanisms that are incentive-compatible in that truthful reporting of values is an equilibrium.

Assuming that traders have general, well-behaved value densities with overlapping supports, MS use the revelation principle to characterize the mechanisms that maximize traders' total expected surplus in the set of incentive-compatible and interim individually rational mechanisms (their Theorems 1 and 2).[9] They then use their characterization to show that in equilibrium, no feasible mechanism can avoid a positive probability of ex post inefficiency (Corollary 1). In the leading case of uniform value densities, they derive a closed-form solution for the mechanism that maximizes traders' total expected surplus subject to incentive constraints (pp. 276-277) and show that it mimics the outcomes of CS's linear double-auction equilibrium. For reference, that equilibrium has ex ante probability of trade 9/32 ≈ 28% and total surplus 9/64 ≈ 0.14, well below the maximum ex post individually rational probability of trade of 50% and surplus 1/6 ≈ 0.17.

[9] Williams (1987) notes that MS's maximization of traders' total expected surplus does not identify all mechanisms associated with outcomes on the incentive-efficient frontier, because the incentive constraints interfere with the transferability of utility that usually follows from quasilinearity. He characterizes the mechanisms associated with all possible welfare weights.

3. A level-*k* model for incomplete-information games

I now define a level-*k* model for incomplete-information games. As in the analysis, I restrict attention to direct mechanisms, for which there is clear evidence to guide the specification.

Recall that a level-*k* player anchors its beliefs in a naive model of other players' responses, *L0,* with which it assesses the payoff implications of its own decisions before thinking about others' responses to 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 usually taken to be uniformly random over the range of feasible decisions. I extend this to games with incomplete information by taking *L0*'s decisions to be uniform over the feasible decisions and *independent of its own value*, following Milgrom and Stokey (1982); Camerer et al. (2004); Crawford and Iriberri (2007); and CKNP (see also Crawford et al., 2013, Section 5).[10] I also assume that a player's level is independent of its value.

Experiments and some analyses of field data (Camerer et al., 2004; Brown, Camerer, and Lovallo, 2012; Brocas et al., 2014) suggest that this generalized level-*k* model gives a reliable, unified account of people's non-equilibrium thinking and their informational naiveté, the often-observed imperfect attention to how others' decisions depend on their private information.[11]

4. A level-*k* analysis of the double auction

I now extend CS's equilibrium analysis of the double auction to Section 3's level-*k* model. I take the population frequency of *L0* and *L3*, *L4*, … to be zero and focus on *L1*s and *L2*s, which predominate empirically and serve to illustrate my main points. In this section, for simplicity, I consider mainly cases in which buyers are matched with sellers of the same level. For *L1*s this section's analysis applies to any value densities; for *L2*s it assumes uniform value densities.

Denote a level-*k* buyer's bidding strategy $b_k(V)$ and a level-*k* seller's asking strategy $a_k(C)$.

[10] Milgrom and Stokey's (1982) notions of Naïve Behavior and First-Order Sophistication, which they suggest might explain zero-sum trading despite their equilibrium-based No Trade ("Groucho Marx") Theorem, are equivalent to an *L1* defined this way and an *L2* best responding to such an *L1*. An *L0* buyer's bid or seller's ask might instead be assumed to be uniformly distributed below (above) its value, thus eliminating weakly dominated bids (asks). However, *L0* represents not a real player but a player's initial, naïve model of others whose values it does not observe. Experiments suggest that few people perform the contingent reasoning a more sophisticated *L0* would require. Another alternative would assume *L0* bids or asks its true value, which is well-defined for games created by direct mechanisms. But a truthful *L0* has less experimental support in this context (Crawford and Iriberri, 2007) and would trivially reduce level-*k* incentive constraints to equilibrium incentive constraints.

[11] Informational naiveté is most famously observed in the winner's curse, which is not relevant with independent private values. However, the kinds of informational naiveté studied here are relevant even with independent private values.

4.1. L1 *buyer and seller*

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 strictly increasing in the relevant range, so the event $a = b$ can be ignored. Solving the first-order condition yields $b_1(V) = 2V/3$, with range [0, 2/3]; and the second-order condition is satisfied. Thus, boundaries aside, an *L1* buyer bids 1/12 more aggressively (that is, bids less) than a linear-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, inducing the buyer to underbid relative to the linear equilibrium.[12]

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 $a_1(C) = 2C/3 + 1/3$, with range [1/3, 1]; and the second-order condition is again satisfied. Boundaries aside, an *L1* seller asks 1/12 more aggressively (that is, asks more) than a linear-equilibrium seller with uniform value densities: An *L1* seller's naïve model of the buyer systematically underestimates the distribution of the buyer's downward-shaded bid, inducing the seller to over-ask relative to the linear equilibrium.

L1 buyers' and sellers' strategies have the same 2/3 slopes as in CS's linear equilibrium with uniform value densities. When *L1*s are matched, trade takes place whenever $V \geq C + 1/2$, so the required value gap is 1/4 larger than in CS's linear equilibrium. The ex ante probability of trade is 1/8 = 12.5% and the total expected surplus is 1/24 ≈ 0.04, far less than the linear equilibrium probability of trade 9/32 ≈ 28% and surplus 9/64 ≈ 0.14, and even further below the maximum individually rational probability of trade 50% and surplus 1/6 ≈ 0.17.

4.2. L2 *buyer and seller*

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

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

[12] Compare Crawford and Iriberri's (2007) analysis of *L1* and *L2* bidding in first-price auctions. Despite the double auction's multiplicity of equilibria, the level-*k* model makes predictions that are generically unique, given traders' levels.

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 zero elsewhere. It thus believes that trade requires $b > 1/3$. For $V \leq 1/3$ it is therefore optimal to bid anything that it believes yields zero probability of trade. In the absence of dominance among such strategies, I set $b_2(V) = V$ for $V \in [0, 1/3]$. For $V > 1/3$, if $g(C)$ is uniform, an *L2* buyer's bid $b_2(V)$ must maximize over $b \in [1/3, 1]$

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

The optimal *L2* strategies are strictly increasing in the relevant range, so the event $a = b$ can again be ignored. Solving the first-order condition $(3/2)(V - \mathrm{b}) - (3/4)(V - 1/3) = 0$ yields $b_2(V) = 2V/3 + 1/9$ for $V \in [1/3, 1]$, with range [1/3, 7/9]; and the second-order condition is satisfied. Boundaries aside, with uniform value densities an *L2* buyer bids 1/36 less aggressively (more) than a linear-equilibrium buyer, and 1/9 less aggressively than an *L1* buyer: An *L2* buyer's model of the seller, though less naïve than an *L1* buyer's, overestimates the distribution of the seller's upward-shaded ask, inducing the buyer to overbid relative to the linear equilibrium.

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

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

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 $f(V)$ 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 that it believes yields 0 probability of trade. In the absence of dominance among such strategies, I set $a_2(C) = C$ for $C \in [2/3, 1]$. For $C < 2/3$, an *L2* seller's ask $a_2(C)$ must maximize over $a \in [0, 2/3]$

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

The second-order condition is satisfied, and the first-order condition $(3/2)(a\text{-}C) + (3/2)(2/3 - C)/2 = 0$ yields $a_2(C) = 2C/3 + 2/9$ for $C \in [0, 2/3]$, with range [2/9, 2/3]. Boundaries aside, with uniform value densities an *L2* seller asks 1/36 less aggressively (less) than a linear equilibrium seller, and 1/9 less aggressively than an *L1* seller: An *L2* seller overestimates the distribution of the buyer's downward-shaded bid, inducing it to under-ask relative to the linear equilibrium.

With uniform value densities, *L2* buyers' and sellers' strategies again have the same 2/3 slopes as in CS's linear equilibrium with uniform densities. When *L2*s are matched, trade takes place whenever $V \geq C + 1/6$, so the required value gap is 1/12 *smaller* than in CS's linear equilibrium. The ex ante probability of trade is 25/72 ≈ 35% and the total expected surplus is 11/72 ≈ 0.15: somewhat more than the linear equilibrium probability 9/32 ≈ 28% and surplus 9/64 ≈ 0.14 and much more than the *L1*s' probability 1/8 = 12.5% and surplus 1/24 ≈ 0.04, but still well below the maximum individually rational probability 50% and surplus 1/6 ≈ 0.17.

4.3. Buyer and seller with mixed levels

When an *L1* buyer and an *L2* seller are matched, or vice versa, trade occurs when $V \geq C + 1/3$, so the needed value gap is 1/12 more than for linear equilibrium traders and 1/6 more than for matched *L2*s, but 1/6 less than for matched *L1*s. With uniform value densities levels affect the gap additively, so with random matching outcomes are determined by population average levels.

5. Level-*k* design when level-*k*-incentive-compatibility is required

This section considers level-*k* design when level-*k*-incentive-compatibility and interim individual rationality are required, continuing to focus on direct mechanisms. Like MS, I take the goal to be maximizing traders' true total expected surplus, but now subject to incentive constraints defined for traders' possibly heterogeneous level-*k* beliefs.

Although in this case level-*k* traders are incentivized to report their own values truthfully, they generally do not expect their partners to report truthfully. Importantly, I assume that even so, traders question neither the mechanism's feasibility nor that the stated mechanism will determine outcomes.[13] I also treat differences in traders' levels as pure differences of opinion (Eliaz and Spiegler, 2008). In particular, traders draw no inferences about each other's levels from the designer's choice of mechanism or, conditionally, from each other's decisions.

I consider three cases: uniform value densities with traders' levels observable or not; general densities with traders' levels observable; and general densities with levels not observable.

[13] If, instead, traders used their beliefs to assess the mechanism's feasibility or credibility, the effect would be to add constraints that for fully responsive mechanisms are generically in conflict with level-*k*-incentive-compatibility. The analysis would then likely resemble Section 5.4's (Lemma 1, Theorem 4) analysis of the case where traders' levels are not observable, in which level-*k* incentive constraints lead to total-surplus-maximizing mechanisms with posted prices and thus limited responsiveness.

5.1 Uniform value densities with traders' levels observable or not

First consider the case of uniform value densities, where MS obtained a closed-form solution for the mechanism that maximizes traders' total expected surplus in the set of equilibrium-incentive-compatible mechanisms, which mimics CS's linear double-auction equilibrium.

Theorem 1 shows that this part of MS's analysis extends exactly to level-*k* models (see also the related results of Gorelkina, 2018, Proposition 1; and de Clippel et al., 2019, Observation 1).

Theorem 1. *With uniform value densities, for any population of level-*k *traders, with levels observable or not, level-*k*-incentive-compatibility and interim individually rationality constraints coincide with equilibrium-incentive-compatibility and interim individually rationality constraints. Thus, MS's mechanism that maximizes traders' total expected surplus in the set of equilibrium-incentive-compatible and interim individually rational mechanisms, also maximizes their total expected surplus in the set of level-*k*-incentive-compatible and interim individually rational mechanisms.*

Proof. The proof is inductive. *L1* traders believe that they face uniform distributions of each other's value reports. When the mechanism is level-*k*-incentive-compatible, with uniform value densities their beliefs are correct. *L1*-incentive-compatibility and interim individual rationality therefore coincide with equilibrium-incentive-compatibility and interim individual rationality. If *L1* traders report truthfully, *L2* traders' beliefs are correct and *L2*-incentive-compatibility and interim individual rationality also coincide with equilibrium-incentive-compatibility and interim individual rationality. And so on, ad infinitum. Like MS's analysis, the level-*k* analysis seeks to maximize traders' true total expected surplus. Thus for any population of level-*k* traders, with levels observable or not, MS's mechanism that maximizes total expected surplus subject to equilibrium incentive constraints also maximizes it subject to level-*k* incentive constraints. ■

*5.2. Level-*k *menu effects and failure of the revelation principle*

Recall that with uniform value densities, the truthful equilibrium of MS's (p. 277) mechanism that maximizes traders' total expected surplus in the set of equilibrium-incentive-compatible and interim individually rational mechanisms is outcome-equivalent to CS's linear double-auction equilibrium. Section 4's analysis shows that no such outcome-equivalence holds for level-*k* models: *L1*s have lower surplus in the double auction than in the truthful equilibrium of MS's mechanism, while *L2*s have higher surplus. Accordingly, Theorem 1 shows only that MS's mechanism maximizes level-*k* traders' total expected surplus subject to level-*k* incentive constraints if it is implemented in its level-*k*- (and equilibrium-) incentive-compatible form.

Outcome-equivalence fails because the two mechanisms have different relationships to *L0*, which influences the correctness of level-*k* beliefs via level-*k* menu effects (CKNP). Using MS's mechanism rectifies *L1*s' beliefs, neutralizing *L1*s' aggressiveness in the double auction and yielding them the same total expected surplus they would have obtained in CS's linear double-auction equilibrium. Using MS's mechanism would also rectify *L2*s' beliefs, but that would neutralize *L2*s' *un*aggressiveness in the double auction, foregoing its benefits. Put another way, if level-*k*-incentive-compatibility were *not* required, as it is in Theorem 1, then *L2*s would obtain higher total surplus in the non-*L2*-incentive-compatible double auction than in MS's mechanism.

Because the revelation principle fails for level-*k* models, it matters whether level-*k*-incentive-compatibility is required. Some analysts argue that incentive-compatibility is essential in applications, albeit mostly in equilibrium analyses where in theory there is no gain from relaxing it (Milgrom, Ausubel, Levin, and Segal, 2012, for auctions; Abdulkadiroglu and Sönmez, 2003, for school choice). Others are willing to consider mechanisms that are not equilibrium-incentive-compatible (Myerson, 1981, for the first-price sealed-bid auction; Erdil and Ergin, 2008, or Abdulkadiroglu, Che, and Yasuda, 2011, for the Boston school-choice mechanism). Here I don't attempt to resolve this issue, which is mainly empirical. Instead I consider both cases in turn.

5.3. General value densities with traders' levels observable

Now consider the case of general well-behaved value densities, with traders' levels observable so that the mechanism can be tailored to their combination of levels.

Assuming ex post expected budget balance as in MS's analysis, the payoff-relevant aspects of a direct mechanism are $p(v, c)$, the probability the object is traded, and the expected money transfer $x(v, c)$, 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 let $g^k(c; p, x)$ and $G^k(c; p, x)$ be 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 $g^2(c;p,x) \equiv g(\alpha_1^{-1}(c;p,x))$. Even higher-level beliefs are defined analogously. I sometimes suppress the dependence of higher-level beliefs on (p, x).

Write the buyer's and seller's expected monetary payments, probabilities of trade, and expected utilities as functions of their value reports *v* and *c* and their true values *V* and *C*:

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

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

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

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

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

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

$$\text{(3)} \qquad U_B^k(V,V) \geq 0 \text{ and } U_S^k(C,C) \geq 0.$$

Theorems 2 and 3 extend MS's (Theorems 1-2) characterization of mechanisms that maximize traders' total expected surplus in the set of equilibrium-incentive-compatible and interim individually rational mechanisms, to level-*k* models with traders' levels observable, showing that in this case MS's characterization is qualitatively robust to level-*k* thinking.[14]

Theorem 2. *Assume that traders' levels are observable, say* i *for the buyer and* j *(possibly ≠ i) for the seller. Then for any mechanism (p, x) that is level-*k*-incentive-compatible for those traders,*

$$\text{(4)} \qquad U_B^i(0,0) + U_S^j(1,1) = \min\nolimits_{V\in[0,1]} U_B^i(V,V) + \min\nolimits_{C\in[0,1]} U_S^j(C,C)$$

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

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

[14] Tilman Börgers and referees noted important errors in my previous versions of Theorem 2 and helped me to correct them.

(5) $0 \leq \int_0^1 \int_0^1 \left\{ \left[V - \frac{1-F(V)}{f(V)} \right] \left[\frac{g^i(C;p,x)}{g(C)} \right] - \left[C + \frac{G(C)}{g(C)} \right] \left[\frac{f^j(V;p,x)}{f(V)} \right] \right\} p(V,C) f(V) g(C) dC dV.$[15]

Proof. The proof adapts MS's (pp. 269-271) proof, with adjustments for traders' level-*k* beliefs. By (2), $P_B^i(\cdot)$ is weakly increasing and $P_S^j(\cdot)$ is weakly decreasing for any given (*p*, *x*). When traders' levels are observable, the mechanism designer has all the information required to incentivize them to reveal their true values. As in MS's proof, (1) and (2) yield necessary and sufficient conditions for traders' level-*k* incentive-compatibility, namely that for all *V* and *C*:

(6) $$U_B^i(V,V) = U_B^i(0,0) + \int_0^V P_B^i(\hat{v}) d\hat{v} = U_B^i(0,0) + \int_0^1 \int_0^V P_B^i(\hat{v}) d\hat{v} f(V) dV$$
$$= U_B^i(0,0) + \int_0^1 [1 - F(V)]\, P_B^i(V) dV = U_B^i(0,0) + \int_0^1 \int_0^1 \{1 - F(V)\} g^i(C) p(V,C) dC dV.$$

and

(7) $$U_S^j(C,C) = U_S^j(1,1) + \int_C^1 P_S^j(\hat{c}) d\hat{c} = U_S^j(1,1) + \int_0^1 \int_C^1 P_S^j(\hat{c}) d\hat{c} g(C) dC$$
$$= U_S^j(1,1) + \int_0^1 G(C)\, P_S^j(C) dC = U_S^j(1,1) + \int_0^1 \int_0^1 G(C) f^j(V) p(V,C) dC dV.$$

(6) and (7) imply that $U_B^i(V,V)$ is increasing and $U_S^j(C,C)$ is decreasing, and show that $U_B^i(0,0) \geq 0$ and $U_S^j(1,1) \geq 0$ suffice for interim individual rationality for all *V* and *C* as in (3).

To derive the incentive budget constraint (5) (my term, not MS's), which is analogous to MS's (p. 269) equilibrium-based constraint (2), note that both the mechanism designer and traders know the mechanism and the true value densities. When traders' levels are observable the designer can use the predictability of level-*k* beliefs to calculate, for any given mechanism, the true expected surpluses required to incentivize them to report truthfully, just as in MS's analysis. With truthful reporting, the required total expected surplus is obtained by taking the expectations of the right-hand sides of (6) and (7) over the *true* value densities and summing as follows:

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

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

[15] With correct, equilibrium beliefs, $g^i(C;p,x) \equiv g(C)$ and $f^j(V;p,x) \equiv f(V)$, and (4) and (5) reduce to MS's equilibrium-based expressions. Because level-*k* beliefs are correct for uniform value densities, that is an alternative proof of Theorem 1.

This required total surplus must not exceed the true total surplus the mechanism yields, which is

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

Equating (9) to the expression on the far right-hand side of (8) and simplifying yields (4). Given that $U_B^i(0{,}0) \geq 0$ and $U_S^j(1{,}1) \geq 0$ suffice for interim individual rationality for all V and *C*, (3) and (4) imply (5). (4) and (5) may appear to "compare apples and oranges" because in general traders' level-*k* beliefs differ from the true value densities. But as the proof shows, those comparisons are valid under my assumptions because traders are incentivized to reveal their true values and they do not use their beliefs to question the mechanism's feasibility or credibility.

Finally, given $p(\cdot,\cdot)$ mapping [0, 1]×[0, 1] into [0, 1], the monotonicity of $P_B^j(\cdot)$ and $P_S^k(\cdot)$ implies that a level-*k* adaptation of MS's (pp. 270-271) transfer function:

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

ensures that (*p*, *x*) is incentive-compatible and interim individually rational for each trader. ■

In MS's analysis the incentive-compatibility constraints can be characterized trader by trader even though they are based on an equilibrium that is a fixed point in which traders interact, because the revelation principle (including the assumed ability to choose among equilibria) decouples traders' problems that determine whether truth-telling is an equilibrium. The level-*k* analysis can rely on analogous methods because level-*k* traders avoid fixed-point reasoning—as experimental subjects normally do (Crawford et al., 2013, Sections 3 and 5)—and the problems that describe their strategic thinking decouple even though the revelation principle fails.

Adapting MS's Theorem 2's statement and proof (pp. 274-276) to traders' level-*k* beliefs, Theorem 3 completes the characterization of mechanisms that maximize traders' total expected surplus in the set of level-*k*-incentive-compatible and interim individually rational mechanisms.

Theorem 3. *Assume that traders' levels are observable, say* i *for the buyer and* j *(possibly ≠ i) for the seller. If there exists a mechanism (p, x) such that* $U_B^i(0{,}0) = U_S^j(1{,}1) = 0$, *the level-*k *incentive budget constraint (5) is satisfied, and the Kuhn-Tucker conditions*

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

$V\left[(1+\lambda)\frac{g^i(C;p,x)}{g(C)}\right] - C\left[(1+\lambda)\frac{f^j(V;p,x)}{f(V)}\right] \leq \lambda\left[\frac{1-F(V)}{f(V)}\frac{g^i(C;p,x)}{g(C)} + \frac{G(C)}{g(C)}\frac{f^j(V;p,x)}{f(V)}\right]$ *when* $p(V, C) = 0$ *and*

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

$V\left[(1+\lambda)\frac{g^i(C;p,x)}{g(C)}\right] - C\left[(1+\lambda)\frac{f^j(V;p,x)}{f(V)}\right] \geq \lambda\left[\frac{1-F(V)}{f(V)}\frac{g^i(C;p,x)}{g(C)} + \frac{G(C)}{g(C)}\frac{f^j(V;p,x)}{f(V)}\right]$ *when* $p(V, C) = 1$ *are satisfied for some* $\lambda \geq 0$, *then that mechanism maximizes traders' total expected surplus among all mechanisms that are level-*k*-incentive-compatible and interim individually rational for trader's levels. Furthermore, if for that (p, x) the function* $\left[V - \frac{1-F(V)}{f(V)}\right]\left[\frac{g^i(C;p,x)}{g(C)}\right] - \left[C + \frac{G(C)}{g(C)}\right]\left[\frac{f^j(V;p,x)}{f(V)}\right]$ *is weakly increasing in V and weakly decreasing in C on [0,1]×[0,1], then such a mechanism must exist.*

Proof. Consider the problem of choosing $p(\cdot, \cdot)$ to maximize traders' total expected surplus subject to $0 \leq p(\cdot, \cdot) \leq 1$ and (5). The problem is analogous to a consumer's budget problem with the trade probabilities $p(V, C)$ analogous to a continuum of linearly priced goods. Form the Lagrangean (for ease of notation, without pricing out the $p(V, C) \leq 1$ constraints):

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

$$\text{(13)} \qquad +\lambda\int_0^1\int_0^1\left\{\left[V - \frac{1-F(V)}{f(V)}\right]\left[\frac{g^i(C;p,x)}{g(C)}\right] - \left[C + \frac{G(C)}{g(C)}\right]\left[\frac{f^j(V;p,x)}{f(V)}\right]\right\}p(V, C)f(V)g(C)dCdV$$

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

The objective function and the constraint are linear in the $p(V, C)$, so the solution will be "bang-bang", with $p(V, C) = 0$ or 1 almost everywhere. The Kuhn-Tucker conditions require $\lambda \geq 0$, (5), (11), and (12). (11) and (12) are analogous to marginal-utility-to-price-ratio first-order conditions determining which of the $p(V, C)$ should be set equal to one.

For level-*k* traders with observable levels, I show below that it is theoretically possible for (5) to be slack at the solution, with the optimal $\lambda = 0$ then, in which case an optimal $p(V, C) = 1$ if and only if $V \geq C$. Normally, however, (5) is binding and the optimal $\lambda > 0$, in which case some of the $p(V, C) = 0$ even if $V > C$. Setting λ so that when (11)-(12) are satisfied and $U_B^i(0,0) =$

$U_S^j(1{,}1) = 0$, (5) holds with equality then yields a mechanism that maximizes traders' total expected surplus among all level-*k*-incentive-compatible and individually rational mechanisms.

Finally, if for that mechanism (p, x), $\left[V - \frac{1-F(V)}{f(V)}\right]\left[\frac{g^i(C;p,x)}{g(C)}\right] - \left[C + \frac{G(C)}{g(C)}\right]\left[\frac{f^j(V;p,x)}{f(V)}\right]$ is weakly increasing in *V* and weakly decreasing in *C* on [0,1]×[0,1], (11)-(12) imply that $P_B^i(\cdot)$ is weakly increasing and $P_S^j(\cdot)$ is weakly decreasing as required in Theorem 2. Continuity and monotonicity arguments like MS's (p. 276) then show that there exists a unique λ > 0 such that (11)-(12) are satisfied and that with $U_B^i(0{,}0) = U_S^j(1{,}1) = 0$, (5) holds with equality. ■

Theorem 3's terms $\left[V - \frac{1-F(V)}{f(V)}\right]\left[\frac{g^i(C;p,x)}{g(C)}\right] - \left[C + \frac{G(C)}{g(C)}\right]\left[\frac{f^j(V;p,x)}{f(V)}\right]$ differ from the virtual values in MS's Theorems 1-2 in that they are adjusted by traders' level-*k* beliefs, so that each term in the difference depends on both traders' values. As a result, Theorems 2 and 3's level-*k* monotonicity conditions may not be satisfied in Myerson's (1981) "regular case" (which rules out true value densities with strong hazard rate variations in the "wrong" direction). Whatever traders' levels, Theorem 2's and 3's monotonicity conditions are not systematically more (or less) stringent than MS's monotonicity conditions. By Theorem 1 they are approximately the same as MS's conditions when the true value densities are close to uniform, for all levels, and this suggests that they will be satisfied whenever traders' level-*k* beliefs are not too extreme.

When traders' levels exceed one, the designer's optimization problem involves a fixed-point recursion for the designer (though not for the traders) because $p(V, C)$ influences the constraints via traders' beliefs, which beliefs in turn influence the optimal $p(V, C)$. Because the continuity of the value densities ensures continuity of the objective and constraint functions and the feasible region is compact, this two-way influence does not interfere with the existence of a solution. Nor is it a conceptual difficulty. However, it does make the problem computationally much more difficult for higher levels, though possibly still tractable via iterative methods.

Comparing the level-*k* incentive budget constraint (5) with MS's constraint (2) (p. 269) and comparing the level-*k* Kuhn-Tucker conditions (11)-(12) with MS's (p. 274) equilibrium-based conditions shows that design features that contribute to maximizing equilibrium traders' total expected surplus subject to incentive constraints, also do so for level-*k* traders, though with different weights. As a result, unless a total-surplus-maximizing mechanism happens to induce traders' level-*k* beliefs that are correct (as with Theorem 1's uniform value densities), the mechanism must involve tacit exploitation of predictably incorrect beliefs ("TEPIB"). TEPIB favors trade at (*V*, *C*) combinations for which level-*k* beliefs make the "prices" (in curly brackets in (11)-(12)) more favorable than they are for equilibrium beliefs: in particular at combinations for which $\frac{f^j(V;p,x)}{f(V)} > 1$ and/or $\frac{g^i(C;p,x)}{g(C)} < 1$, in which level-*k* traders underestimate the likelihood of trade given their values. Moreover, for levels higher than *L1* TEPIB also favors mechanisms that increase the total expected surplus from such trades.

By contrast with MS's result that no equilibrium-incentive-compatible mechanism can assure ex post efficient trade with probability one, for level-*k* traders with observable levels it is theoretically possible that the mechanism that maximizes their total expected surplus subject to the level-*k* incentive constraints *does* assure ex post efficient trade with probability one. To see this, adapt MS's proof of their Corollary 1 (pp. 271-273) for a level-*i* buyer and a level-*j* seller. With value densities supported on [0, 1] and $p(V,C) \equiv 1$ if and only if $V \geq C$, (5) reduces to:

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

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

(14) $= \int_0^1 [Vf(V) + F(V) - 1][G^i(V;p,x) - G^i(0;p,x)] dV - \int_0^1 VG(V) f^j(V;p,x)] dV$

$$= -\int_0^1 [1 - F(V)] G^i(V;p,x) dV + \int_0^1 Vf(V) G^i(V;p,x) dV - \int_0^1 VG(V) f^j(V;p,x) dV.$$

With equilibrium beliefs, the last two terms on the last line of (14) exactly offset each other, leaving only the first term as in MS's proof, which proves their Corollary 1. With level-*k* beliefs, however, the middle term can outweigh the others in some cases. If, for instance, $G^i(V;p,x)$ increases sharply from 0 to 1 near *V* = 1, a calculation shows that the right-hand side of (14) is strictly positive. But the computations for *L1*s reported below suggest that such cases are rare.

MS's Corollary 1 shows that with equilibrium beliefs, the optimal λ > 0 and $p(V,C) = 1$ only if

(15) $$V - C \geq \frac{\lambda}{1+\lambda} \left[\frac{1-F(V)}{f(V)} + \frac{G(C)}{g(C)} \right].$$

Thus, optimal mechanisms for equilibrium traders involve ex post inefficiency for some value combinations, but only in the form of lost opportunities to trade when *V* > *C*. By contrast, (11)-(12) show that for level-*k* traders, when the optimal λ > 0, there is a wedge between $V - C$ and the criterion for $p(V, C) = 1$. The criterion still responds positively to *V* and negatively to *C*, but now with weights that do not exclusively favor high values of $V - C$. As a result, for some true value densities it is possible for the total-surplus-maximizing mechanism to have $p(V, C) = 1$ for some value combinations with $V < C$, allowing a form of ex post-inefficient trade that does not arise in an equilibrium-based analysis. Recall that the rationale for MS's requirement of interim but not necessarily ex post individual rationality is that traders must commit to participate in the mechanism at the interim stage, so that the mechanism can enforce such trades. (11)-(12) show that commitments to such trades can repay their immediate cost, by easing the level-*k* incentive constraints enough to enable trade for more value combinations with $V > C$. Figure 1's examples (cases "0.25, 1.5" and "0.25, 1.75") show that such commitments are a real theoretical possibility even for *L1*s, but that they may be limited to value combinations where *V* and *C* are extreme.

As in MS's analysis, closed-form solutions are available only for the case of uniform value densities, but that case is somewhat unrepresentative because it induces level-*k* beliefs that are correct, so TEPIB can have no influence. Figure 1 illustrates the relationship between equilibrium and level-*k* design for a population of *L1* traders with observable levels for a representative subset of combinations of linear value densities.[16] Specifically, Figure 1 depicts computed trading regions for total-expected-surplus-maximizing mechanisms in the sets of equilibrium- and *L1*-incentive-compatible and interim individually rational mechanisms.[17]

[insert Figure 1 about here]

In Figure 1's cases, total-expected-surplus-maximizing mechanisms for *L1* traders are qualitatively similar to those for equilibrium traders. For density combinations in which *L1* sellers' uniform beliefs underestimate buyers' true densities (solid density curve upward-sloping), the optimal mechanisms for *L1* traders exploit TEPIB to implement trading regions that are supersets of—thus with higher total expected surplus than—those for equilibrium traders (with one exception, case "0.75, 1.75", in which the trading regions overlap slightly). As already noted, in two such combinations with extreme densities, cases "0.25, 1.5" and "0.25, 1.75", the

[16] The computations are feasible for *L1*s, but for *L2*s, 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) and (12)-(13) depend on the transfer function $x(\cdot, \cdot)$ as well as on $p(\cdot, \cdot)$, making the dimensionality of search too high at present. Figure 1* in the online appendix depicts the analogous trading regions for a comprehensive coarse subset of all possible combinations of linear value densities, excluding only extreme combinations that violate the monotonicity conditions for optimality (Theorems 2-3), and including those in Figure 1. The appendix also includes MATLAB code for the computations, developed by Rustu Duran, International School of Economics, Kazakh-British Technical University, Almaty, Kazakhstan.

[17] Note that buyers and sellers are *not* symmetric: The seller's initial ownership of the object breaks the symmetry between them. Even optimal trading regions for equilibrium traders are asymmetric when interchanging the buyer's and the seller's densities.

optimal mechanisms for *L1* traders allow ex post-inefficient trade (*V* < *C*, trading to the right of the main diagonal) for very high values of *V* and *C*. By contrast, for density combinations in which *L1* sellers' uniform beliefs overestimate buyers' (solid density curve downward-sloping) true densities, the optimal trading regions for equilibrium traders are supersets of—with higher total expected surplus than—the optimal trading regions for *L1* traders.

5.4. General value densities with traders' levels not observable

Now consider the case of general well-behaved value densities, assuming that traders' levels are not observable and that the population level distributions for buyers and sellers both include all levels from *L1* up to at least some $K > 1$. The proof of Theorem 1 yields a result that is also useful for the case of general value densities (see also de Clippel et al., 2019, Observation 2).

Lemma 1. *A direct mechanism is level-*k*-incentive-compatible and interim individually rational for all levels from* L1 *up to at least some* K > *1 if and only if it is both* L1- *and equilibrium-incentive-compatible and interim individually rational.*

Proof. The proof is inductive. For general well-behaved value densities, if the mechanism is *L1*-incentive-compatible and interim individually rational, then *L2*s' beliefs are correct and *L2*-incentive-compatibility and interim individual rationality coincide with equilibrium-incentive-compatibility and interim individual rationality. Conversely, if the mechanism is both *L1*- and equilibrium-incentive-compatible and interim individually rational, then it must also be *L2*-incentive-compatible and interim individually rational; and so on ad infinitum. ■

A random posted-price mechanism (Hagerty and Rogerson, 1987; Čopič and Ponsati, 2008, 2016) is a distribution over posted prices π and a probability density $\mu(\cdot)$ such that traders make their value reports after the posted price is drawn and trade occurs at price π with probability $\mu(\pi)$ if $c \leq \pi \leq v$, 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. A random or deterministic posted-price mechanism makes it a weakly dominant strategy for buyers and sellers of all levels to report their true values by making $U_B^1(V, v)$ and $U_B^*(V, v)$ locally independent of v unless $V = \pi$, with discontinuities there that are consistent with the global optimality of $v = V$; and by making

$U_S^1(C,c)$ and $U_S^*(C,c)$ locally independent of c unless $C = \pi$, with discontinuities there that are consistent with the global optimality of $c = C$.

Theorem 4 shows that when traders' levels are not observable and the population level distributions for buyers and sellers include all levels from *L1* up to at least some *K* > 1, then with the exception of certain nongeneric traders' value densities, a mechanism that maximizes traders' total expected surplus among all level-*k*-incentive-compatible and interim individually rational mechanisms must be a posted-price mechanism with a particular price.[18]

Recall that Theorem 1 shows that with uniform value densities (even with multiple, unobserved levels) MS's mechanism that maximizes traders' total expected surplus subject to equilibrium incentive constraints also maximizes traders' total expected surplus subject to level-*k* incentive constraints. Because MS's mechanism is not a posted-price mechanism, that would contradict Theorem 4's conclusion. The theorem therefore requires that $f(\cdot)$, g$(\cdot) \neq 1$ almost everywhere. Even with nonuniform value densities, there may be step-function equilibria of the double auction whose bid and ask distributions have discrete supports that happen to coincide with values where the distribution functions mimic uniform distributions, which is all that matters for traders' best responses in such equilibria (Leininger, Linhart, and Radner, 1989, Section 3.4). Because mechanisms based on such equilibria would also be incentive-compatible for level-*k* traders but are not posted-price mechanisms, they too would contradict Theorem 4. The theorem therefore requires that there is no *x* for which both $F(x) = x$ and $G(x) = x$. The proof suggests that there may be other exceptional cases, which are also nongeneric.

Theorem 4. *Assume that traders' levels are not observable and that the population level distributions for buyers and sellers include all levels* from L1 *up to at least some* K > *1. Assume further that* $f(\cdot)$, g$(\cdot) \neq 1$ *almost everywhere and that there is no x for which both* $F(x) = x$ and $G(x) = x$. *Then, with the possible exception of certain other nongeneric value densities, a mechanism that maximizes traders' total expected* surplus *among all level*-k-*incentive-compatible and interim individually rational mechanisms is equivalent to a deterministic posted-price mechanism with* $U_B^i(0,0) = U_S^j(1,1) = 0$ *for all levels* i *and* j *in the populations and an optimal posted price π that satisfies the first-order condition:*

[18] Larry Samuelson and Rene Saran noted errors in previous versions of Theorem 4 and made suggestions that led to this proof.

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

Proof. Recall (1) and (2) for *L1* traders and their analogues (with * superscripts below) for equilibrium traders. By Lemma 1, level-*k*-incentive-compatibility for all levels holds if and only if, for all *V* and *C*:

(17) $$U_B^1(V,V) \geq U_B^1(V,v) = VP_B^1(v) - X_B^1(v) = V\int_0^1 p(v,\hat{c})d\hat{c} - \int_0^1 x(v,\hat{c})d\hat{c},$$

(18) $$U_B^*(V,V) \geq U_B^*(V,v) = VP_B^*(v) - X_B^*(v) = V\int_0^1 p(v,\hat{c})g(\hat{c})d\hat{c} - \int_0^1 x(v,\hat{c})g(\hat{c})d\hat{c},$$

(19) $$U_S^1(C,C) \geq U_S^1(C,c) = X_S^1(c) - CP_S^1(c) = \int_0^1 x(\hat{v},c)d\hat{v} - C\int_0^1 p(\hat{v},c)d\hat{v},$$ and

(20) $$U_S^*(C,C) \geq U_S^*(C,c) = X_S^*(c) - CP_S^*(c) = \int_0^1 x(\hat{v},c)f(\hat{v})d\hat{v} - C\int_0^1 p(\hat{v},c)f(\hat{v})d\hat{v}.$$

Standard arguments based on (6) and (7) and the analogous expressions for equilibrium traders show that $U_B^1(V,V)$, $U_B^*(V,V)$, $U_S^1(C,C)$, and $U_S^*(C,C)$ are almost everywhere differentiable in *V* or *C*. Because $f(\cdot)$ and g(·) are continuous and the theorem's conditions on *F*(·) and *G*(·) rule out mechanisms based on Leininger et al.'s step-function equilibria (in which truthful reporting would make *F*(·) and *G*(·) effectively discontinuous at the supports of the equilibrium mixed strategies —abusing notation to avoid introducing new notation for the "mechanisms based on"), differentiability of $U_B^1(V,V)$, $U_B^*(V,V)$, $U_S^1(C,C)$, and $U_S^*(C,C)$ can fail only where $p(v,c)$ is discontinuous. Therefore, $p(v,c)$ is almost everywhere continuous.

If, away from its points of discontinuity, $p(v,c)$ is not locally constant in *v* and *c*, it creates locally smooth tradeoffs between *v* and *c* along its level curves. Differentiating (17)-(20) shows that such tradeoffs generically make (17) and (18) inconsistent and (19) and (20) inconsistent.

Absent such local tradeoffs, $p(v,c)$ is locally constant in *v* and *c* away from its points of discontinuity. Given its weak monotonicity, the locations of the discontinuities are described by a step function relating *v* to *c* (or vice versa), with at most a countable number of steps. If there is more than one step for any given *c*, traders on the boundaries between them must be indifferent. But their indifference conditions are those that would hold in a Leininger et al. step-function equilibrium, and are thus ruled out by the theorem's conditions on *F*(·) and *G*(·). Thus, in any level-*k*-incentive-compatible mechanism, the step function has exactly one step for any given *c*.

Because the $p(v,c)$ enter both the design problem's objective function and constraints linearly, we can take $p(v,c) = 0$ or 1 almost everywhere without loss of generality. The incentive-compatibility constraints (17)-(20) imply that $x(v,c)$ is constant when $p(v,c) = 0$, and is also constant when $p(v,c) = 1$. By an argument like Theorem 3's, it is optimal to set

$$(21) \qquad U_B^1(0,0) = U_B^*(0,0) = U_S^1(1,1) = U_S^*(1,1) = 0,$$

which, given Lemma 1, holds for all levels i and j. Thus, $x(v,c) = 0$ whenever $p(v,c) = 0$.

Finally, a total expected surplus-maximizing deterministic posted-price mechanism would choose the posted price π to solve:

$$(22) \qquad \max_{\{0 \le \pi \le 1\}} \int_\pi^1 \int_0^\pi (V - C)\, g(C) f(V) dC dV.$$

(15) gives the first-order condition for this problem. Multiple optima appear to be possible, but randomizing serves no purpose and a deterministic posted price is always among the optima. ■

Corollary 1. *Assume that level-*k *traders' levels are not observable. If traders have positive value densities with overlapping supports, then no level-*k*-incentive-compatible and interim individually rational mechanism can assure ex post efficiency with probability one.*

Proof. The proof is immediate as MS's Corollary 1 applies in Theorem 4's exceptional cases. ■

A posted price foregoes the sensitive dependence on reported values of the mechanisms that maximize traders' total expected surplus when their levels are observable (Section 5.3, Theorems 2-3). Even though Theorem 1 does not apply directly to Theorem 4's case of general value densities, it allows an estimate of the cost of giving up such sensitivity when the value densities are close to uniform and traders' levels are not observable. With exactly uniform densities, the surplus-maximizing mechanism yields probability of trade 9/32 ≈ 28% and total expected surplus 9/64 ≈ 0.14 (Section 5.1). With levels unobservable and uniform densities, the optimal posted price is ½, which yields probability of trade 1/4 = 25% and surplus 1/8 = 0.125. The optimal posted price with approximately uniform densities yields approximately these results, a modest cost for the mechanism's robustness to the unpredictability of traders' strategic thinking.

That the surplus-maximizing mechanism when levels are not observable makes truthful reporting a dominant strategy for all levels is an alternative, behaviorally plausible rationale for the dominant-strategy implementation often assumed in robust mechanism design (Hagerty and Rogerson, 1987; Čopič and Ponsatí, 2008, 2016). When levels are not observable the surplus-maximizing mechanism comes close to satisfying Wilson’s (1987) desiderata in that its *rules* are distribution-free. However, the optimal posted price is sensitive to the value densities in (16). Even so, the optimal price can be implemented without knowing the densities, fully satisfying Wilson’s desiderata, via Čopič and Ponsatí’s (2008) dynamic, continuous-time double auction, in which the auctioneer reveals traders’ bids only when they become compatible.

6. Level-*k* design when level-*k*-incentive-compatibility is not required

This section records some observations about design when level-*k*-incentive-compatibility is not required, allowing direct mechanisms that create incentives to lie but continuing to assume that traders best respond to them. One can still define a general class of feasible direct mechanisms, with payoff-relevant outcomes $p(v, c)$ and $x(v, c)$. But their effects can no longer be tractably captured via incentive constraints and must be modeled directly via traders’ level-*k* responses. I consider the cases where traders’ levels are observable and not observable in turn.

6.1. *Traders’ levels observable*

With traders’ levels observable, assume uniform value densities for simplicity and consider double auctions with reserve prices, as a proxy for what is achievable via direct mechanisms. Reserve prices have no benefits if level-*k* traders anchor beliefs on an *L0* that is uniformly random on the full range of values [0, 1]. However, a double auction with a restricted menu of bids or asks might make level-*k* traders anchor on the restricted menu instead of [0, 1]. I know of no evidence for such an *L0* specification, but analogous menu effects are commonplace in marketing. Such anchoring can make reserve prices useful for level-*k* traders: CKNP showed that in first-price auctions with level-*k* bidders such anchoring can change the optimal reserve price, often yielding the seller expected revenue that exceeds Myerson’s (1981) bound.

In bilateral trading via double auction without reserve prices, *L1* traders believe they face bids or asks uniformly distributed on [0, 1], yielding outcomes that do not maximize total expected surplus subject to *L1* incentive constraints. In a double auction with reserve prices for buyer’s bids of ¾ and seller’s asks of ¼, if *L1* traders anchor on the restricted menu, they bid or ask as if facing asks or bids uniformly distributed on [¼, 1] or [0, ¾] respectively, or equivalently (given the ranges of their optimal bids or asks) on [¼, ¾] for both: exactly the ranges of serious bids or asks in CS’s linear double-auction equilibrium (Section 2). Thus, a double auction with those reserve prices rectifies *L1* traders’ beliefs and is outcome-equivalent to

MS's mechanism that maximizes total expected surplus for equilibrium traders. The probability of trade is 9/32 ≈ 28% and the total surplus is 9/64 ≈ 0.14 (Section 2.2), far higher than matched *L1*s' probability of trade 1/8 = 12.5% and total surplus 1/24 ≈ 0.04 in the double auction without reserve prices (Section 4.1). Pushing the reserve prices beyond ¾ and ¼ further reduces the value gap needed for trade, which is a benefit, other things equal; but it also precludes some bids or asks that would allow trade. The cost of precluding those bids or asks exceeds the benefits, and computations not reproduced here show that reserve prices of ¾ and ¼ are in fact optimal.

With traders' levels observable and uniform value densities, for *L2*s a double auction without reserve prices already improves upon MS's mechanism that maximizes total expected surplus for equilibrium traders, or a mechanism that maximizes total surplus in the set of *L2*-incentive-compatible and interim individually rational mechanisms (Sections 2.2, 4.2, and 5.1). Feasible reserve prices (restricted to [0, 1]) bring *L2*s' beliefs closer to equilibrium beliefs, reducing the unaggressiveness that allows the double auction without reserve prices to yield better outcomes for them. Computations not reproduced here show that a double auction without reserve prices is in fact optimal. It has probability of trade 25/72 ≈ 35% and total surplus 11/72 ≈ 0.15, higher than the equilibrium probability of trade 9/32 ≈ 28% and surplus 9/64 ≈ 0.14 (Section 4.2).

6.2. *Traders' levels not observable*

With traders' levels not observable, one can estimate the potential benefits of allowing direct, non-level-*k*-incentive-compatible mechanisms. Suppose for example that the population is known to include a high frequency of one particular level with low frequencies of one or more other levels. With multiple, unobservable levels and nonuniform value densities, requiring level-*k*-incentive-compatibility significantly lowers total expected surplus (Section 5.4, Theorem 4). If the level frequencies are extreme enough, a mechanism that maximizes total surplus in the set of level-*k*-incentive-compatible and interim individually rational mechanisms for only the high-frequency level (Section 5.3, Theorems 2-3), or possibly a non-level-*k*-incentive-compatible mechanism (Section 6.1), will yield more total surplus than a mechanism that maximizes total surplus in the set of level-*k*-incentive-compatible mechanisms for all levels in the population.

Assuming approximately uniform value densities allows an estimate of the cost of requiring level-*k*-incentive-compatibility for all levels in such cases, even though with uniform densities the cost is zero by Theorem 1. With uniform densities the optimal posted-price mechanism has optimal price ½, probability of trade 1/4 = 25%, and total expected surplus 1/8 = 0.125, all independent of the population level frequencies (Section 5.4, Theorem 4). With approximately uniform densities the optimal posted-price mechanism will yield approximately those results. By contrast, with uniform densities the mechanism that maximizes total expected surplus subject only to *L1*-incentive-compatibility constraints yields probability of trade 9/32 ≈ 28% and expected total surplus 9/64 ≈ 0.14 (Sections 4.1 and 5.1). With approximately uniform densities and a frequency of *L1*s close to one, it will yield approximately these significantly better results.

Similarly, if *L2*'s frequency is high enough, the double auction without reserve prices yields probability of trade 25/72 ≈ 35% and surplus 11/72 ≈ 0.15, an even better result (Section 4.2).

As these examples make clear, insisting on exact implementation may be quite costly with level-*k* traders. And whether or not traders' levels are observable, relaxing level-*k*-incentive-compatibility when an application does not require it can yield total expected surplus-maximizing mechanisms that differ qualitatively as well as quantitatively from those that maximize surplus for equilibrium traders, possibly with a significant increase in surplus.

7. **Related literature**

Beyond CS's and MS's analyses, this paper builds on Crawford and Iriberri's (2007) positive level-*k* analysis of auctions and CKNP's level-*k* analysis of optimal independent-private-value auctions, which builds on Myerson's (1981) equilibrium analysis of optimal auctions.

This paper's closest relatives in the recent literature are de Clippel et al. (2019) and Kneeland (2018). Both study level-*k* implementation of social choice rules, allowing general distributions of unobservable levels, and allowing both direct mechanisms and indirect mechanisms in which players report their levels as well as their private information.

Kneeland assumes uniform random *L0*s and allows mechanisms to treat levels unequally, as here; and considers both single- and set-valued rules. She gives 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 equal treatment of levels and consider a range of *L0* specifications.[19] Their main result (Theorem 1) allows *L0* to be chosen as part of the implementation. Like Kneeland, they show under otherwise mild restrictions that for single-valued rules, robust level-*k* and equilibrium incentive constraints coincide. Requiring single-valued rules 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 Theorem 4, which shows that a posted-price mechanism is then generically optimal. That result would also hold for de Clippel et al.'s wider set of *L0* anchors, but treating them as exogenous, as here and as the experimental evidence suggests. Generalizing Theorem 4 would also make equal treatment of levels a conclusion rather than an assumption. However, my results for unobservable levels, unlike theirs, are limited to the bilateral trading environment.

De Clippel et al.'s and Kneeland's analyses have no counterparts to my analyses of cases where traders' levels are observable, with or without requiring level-*k* incentive-compatibility.

In other work on nonequilibrium design, Hagerty and Rogerson (1987), Bulow and 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. Börgers and Li (2019) characterize mechanisms that are "strategically simple" in the sense that players need form only first-order beliefs, with applications to voting and bilateral trade. Gorelkina (2018) conducts a level-*k* analysis of the expected-externality mechanism.

In more abstract settings, Mookherjee and Reichelstein (1992) study dominant-strategy implementation; Matsushima (2007, 2008) studies implementation via finitely iterated dominance; and Bergemann and Morris (2009) and Bergemann, Morris, and Tercieux (2011)

[19] 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.

study implementation in rationalizable strategies. Ollár and Penta (2017, 2019), study rationalizable implementation via direct mechanisms under varying degrees of robustness and provide sensitivity results that are closely connected to level-*k* implementation.[20]

8. Conclusion

This paper has revisited MS's 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 anchoring of level-*k* beliefs on *L0* creates menu effects that make the revelation principle fail, so that neither restricting attention to direct mechanisms nor imposing level-*k*-incentive-compatibility are without loss of generality. If one nonetheless focuses on direct mechanisms, the characterization of mechanisms that maximize total expected surplus depends on two things: whether only level-*k*-incentive-compatible and interim individually rational mechanisms are truly feasible and whether traders' levels can be observed.

My main results are for cases where only level-*k*-incentive-compatible and interim individually rational mechanisms are truly feasible. Then, if traders' value densities are uniform, MS's equilibrium-based result that the incentive-compatible mechanism that mimics CS's linear double-auction equilibrium also maximizes traders' total expected surplus subject to incentive constraints, generalizes exactly to level-*k* models with any distribution of levels, observable or not. However, with level-*k* traders the mechanism must be implemented not as the double auction but in its level-*k*-incentive-compatible form, a failure of the revelation principle.

If, instead, only level-*k*-incentive-compatible and interim individually rational mechanisms are feasible, traders' levels are observable, and they have general well-behaved value densities, then MS's characterization of mechanisms that maximize traders' total expected surplus subject to incentive constraints generalizes qualitatively to level-*k* models, with a novel feature, tacit exploitation of predictably incorrect beliefs. MS's result that in equilibrium no mechanism can assure ex post efficient trade with probability one does not quite generalize to level-*k* models.

If only level-*k*-incentive-compatible and interim individually rational mechanisms are feasible but traders' levels are not observable, with general well-behaved value densities,

[20] Glazer and Rubinstein (1998), Neeman (2003), Eliaz and Spiegler (2006, 2007, 2008), and Wolitzky (2016) study design when the "behavioral" aspect concerns individual decisions or judgment. Bartling and Netzer (2016) and Bierbrauer and Netzer (2016) study robustness to various kinds of social preferences in auction design and implementation of social choice rules.

generically a posted price mechanism, which makes truthful revelation of values a weakly dominant strategy and thereby incentivizes all levels despite their differences in beliefs, maximizes traders' total expected surplus subject to level-*k* incentive constraints. The optimal posted price is sensitive to traders' value densities, but it can be implemented dynamically via a device of Čopič and Ponsatí (2008), in a way that satisfies Wilson's (1987) desiderata.

Finally, if direct, non-level-*k*-incentive-compatible mechanisms are feasible and traders best respond to them, total expected-surplus-maximizing mechanisms may take quite different forms. Further study of this case might identify important roles for indirect and/or non-incentive compatible mechanisms, which are irrelevant in an equilibrium analysis by revelation principle fiat. Preliminary analyses show that nonequilibrium design must go beyond weakening the behavioral assumptions under which desirable outcomes can be implemented in equilibrium.

At first glance, MS's analysis appears to depend on the full strength of their assumption that traders will play the desired equilibrium in any game that the designer's choice of mechanism creates. This bundles four distinct behavioral assumptions: decision-theoretic rationality, homogeneity of strategic thinking, and predictability and coordination/correctness of traders' beliefs. The level-*k* analysis unbundles those assumptions, retaining decision-theoretic rationality while relaxing the behaviorally more questionable assumptions of homogeneity and coordination/correctness of beliefs, in a structured way. It also links the predictability of traders' beliefs to the observability of their levels, relaxing predictability in a structured way when levels are not observable and tracing the need for robust implementation to the unpredictability of traders' strategic thinking, a plausible rationale for robustness. Thus, the structure of level-*k* models allows an analysis of the cases where only direct, level-*k*-incentive-compatible mechanisms are feasible with most of the power and precision of MS's equilibrium analysis, retaining its assumptions of decision-theoretic rationality and predictability or partial predictability of traders' beliefs, while dispensing with the equilibrium analysis's strong assumptions of homogeneity of strategic thinking and coordination/correctness of beliefs.

To sum up, the level-*k* model adds generality to MS's analysis of design for bilateral trading in behaviorally important ways while making specific, empirically disciplined predictions that allow an analysis with power and precision close to that of their equilibrium analysis. I hope that this analysis will encourage further study of design for structural nonequilibrium models.

References

Abdulkadiroglu, Atila, Che, Yeon-Koo, Yasuda, Yosuke, 2011. Resolving conflicting preferences in school choice: the “Boston Mechanism” reconsidered. American Economic Review 101, 1–14.

Abdulkadiroglu, Atila, Sönmez, Tayfun, 2003. School choice: a mechanism design approach. American Economic Review 93, 729-747.

Agranov, Marina, Potamites, Elizabeth, Schotter, Andrew, Tergiman, Chloe, 2012. Beliefs and endogenous cognitive levels: an experimental study. Games and Economic Behavior 75, 449-463.

Alaoui, Larbi, Penta, Antonio, 2016. Endogenous depth of reasoning. Review of Economic Studies 83, 1297-1333.

Alaoui, Larbi, Janezic, Katharina A., Penta, Antonio, 2020. Reasoning about other's reasoning. Journal of Economic Theory 189, 105091.

Aumann, Robert J., Brandenburger, Adam, 1995. Epistemic conditions for Nash equilibrium. Econometrica 63, 1161-1180.

Bartling, Björn, Netzer, Nick, 2016. An externality-robust auction: theory and experimental evidence. Games and Economic Behavior 97, 185-204.

Bergemann, Dirk, Morris, Stephen, 2009. Robust implementation in direct mechanisms. Review of Economic Studies,76, 1175-1204.

Bergemann, Dirk, Morris, Stephen, Tercieux, Olivier, 2011. Rationalizable implementation. Journal of Economic Theory 146, 1253-1274.

Bernheim, B. Douglas, 1984. Rationalizable strategic behavior. Econometrica 52, 1007-1028.

Bierbrauer, Felix, Netzer, Nick, 2016. Mechanism design and intentions. Journal of Economic Theory 163, 557-603.

Börgers, Tilman, Li, Jiangtao, 2019. Strategically simple mechanisms. Econometrica 87, 2003-2035.

Brocas, Isabelle, Carrillo, Juan D., Camerer, Colin F., Wang, Stephanie W., 2014. Imperfect choice or imperfect attention? Understanding strategic thinking in private information games. Review of Economic Studies 81, 944-970.

Brown, Alexander, Camerer, Colin F., Lovallo, Dan, 2012. To review or not to review? Limited strategic thinking at the movie box office. American Economic Journal: Microeconomics 4,1-28.

Bulow, Jeremy, Roberts, John, 1989. The simple economics of optimal auctions. Journal of Political Economy 97, 1060-1090.

Camerer, Colin F., Ho, Teck-Hua, Chong, Juin Kuan, 2004. A cognitive hierarchy model of games. Quarterly Journal of Economics 119, 861-898.

Chatterjee, Kalyan, Samuelson, William, 1983. Bargaining under incomplete information. Operations Research 31, 835-851.

Chen, Yan, Ledyard, John O., 2008. Mechanism design experiments. In: Durlauf, Steven, Blume, Lawrence (Eds.), The New Palgrave Dictionary of Economics, 2nd Edition, Macmillan, London.

Čopič, Jernej, Ponsatí, Clara, 2008. Robust bilateral trade and mediated bargaining. Journal of the European Economic Association 6, 570-589.

Čopič, Jernej, Ponsatí, Clara, 2016. Optimal robust bilateral trade: Risk neutrality. Journal of Economic Theory 163, 276–287.

Costa-Gomes, Miguel A., Crawford, Vincent P., 2006. Cognition and behavior in two-person guessing games: An experimental study. American Economic Review 96, 1737-1768.

Crawford, Vincent P., Costa-Gomes, Miguel A., Iriberri, Nagore, 2013. Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. Journal of Economic Literature 51, 5-62.

Crawford, Vincent P., Iriberri, Nagore, 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., Kugler, Tamar, Neeman, Zvika, Pauzner, Ady, 2009. Behaviorally optimal auction design: An example and some observations. Journal of the European Economic Association 7, 365-376.

de Clippel, Geoffroy, Saran, Rene, Serrano, Roberto, 2019. Level-*k* mechanism design. Review of Economic Studies 86, 1207-1227.

Eliaz, Kfir, Spiegler, Ran, 2006. Contracting with diversely naïve agents. Review of Economic Studies 73, 689–714.

Eliaz, Kfir, Spiegler, Ran, 2007. A mechanism-design approach to speculative trade. Econometrica 75, 875–884.

Eliaz, Kfir, Spiegler, Ran, 2008. Consumer optimism and price discrimination. Theoretical Economics 3, 459–497.

Erdil, Aytek, Ergin, Haluk, 2008. What's the matter with tie-breaking? Improving efficiency in school choice. American Economic Review 98, 669-689.

Glazer, Jacob, Rubinstein, Ariel, 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., Rogerson, William P., 1987. Robust trading mechanisms. Journal of Economic Theory 42, 94–107.

Katok, Elena, Sefton, Martin, Yavas, Abdullah, 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, University College London. http://www.tkneeland.com/uploads/9/5/4/8/95483354/levelk_mechanismdesign_24.10.2018.pdf

Leininger, Wolfgang, Linhart, Peter B., Radner, Roy, 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, Ausubel, Lawrence, Levin, Jon, Segal, Ilya, 2012. Incentive auction rules option and discussion. Rreport to the Federal Communications Commission by Auctionomics and Power Auctions; https://apps.fcc.gov/edocs_public/attachmatch/FCC-12-118A2.pdf

Milgrom, Paul R., Stokey, Nancy, 1982. Information, trade and common knowledge. Journal of Economic Theory 26, 17-26.

Milgrom, Paul, Weber, Robert, 1982. A theory of auctions and competitive bidding. Econometrica 50, 1089-1122.

Mookherjee, Dilip, Reichelstein, Stefan, 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., Satterthwaite, Mark A., 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.

Ollár, Mariann, Penta, Antonio, 2017. Full implementation and belief restrictions. American Economic Review 107, 2243-2277.

Ollár, Mariann, Penta, Antonio, 2019. Implementation via transfers with identical but unknown distributions. Working Paper 1126, Barcelona Graduate School of Economics

Pathak, Parag, 2017. What really matters in designing school choice mechanisms. In Honore, Bo, Pakes, Ariel, Piazessi, Monika, Samuelson, Larry (Eds.), Advances in Economics and Econometrics, 11th World Congress of the Econometric Society, Cambridge University Press, Cambridge, UK.

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.

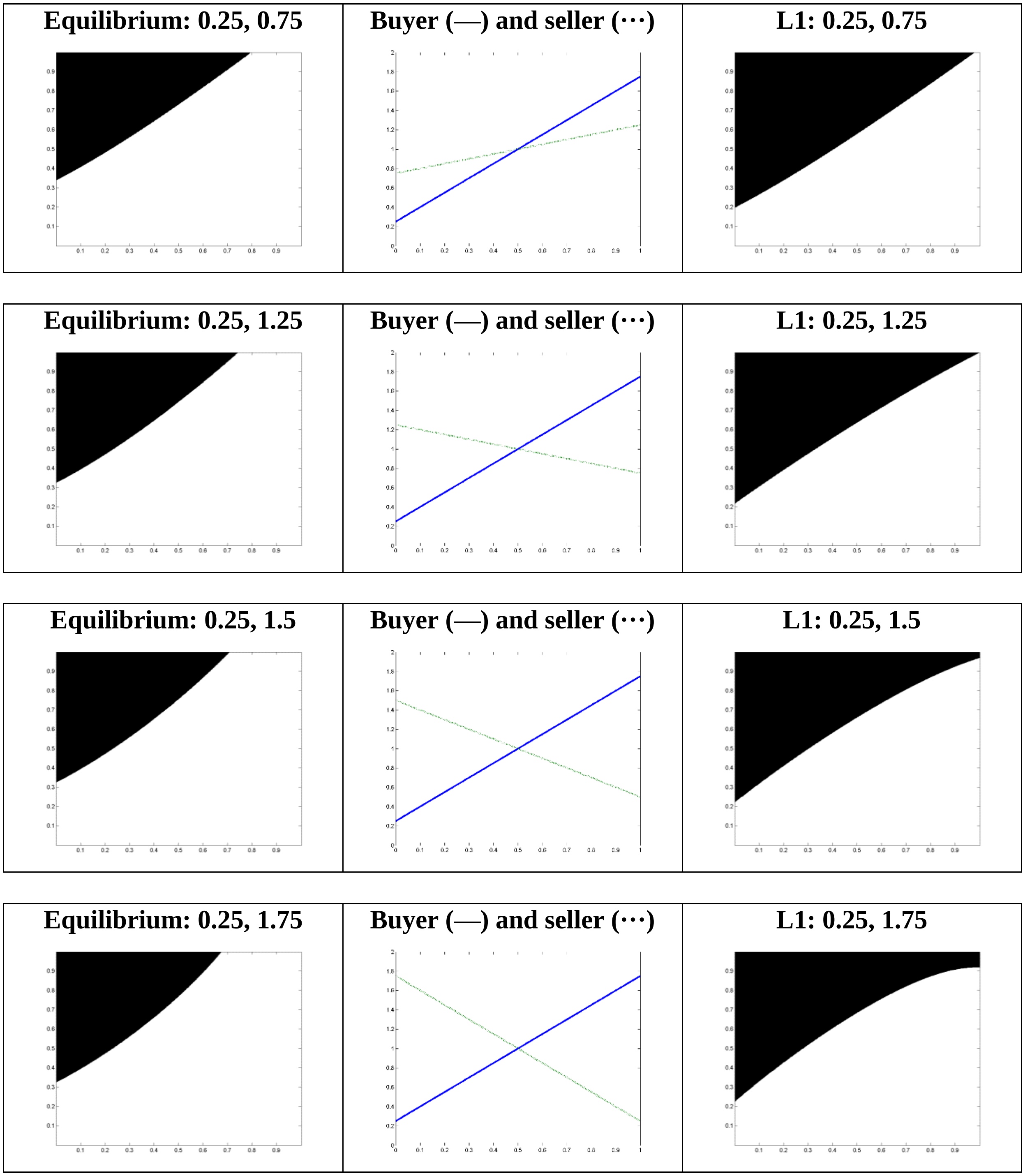
Satterthwaite, Mark, Williams, Steven R., 1989. Bilateral trade with the sealed bid k-double auction: Existence and efficiency. *Journal of Economic Theory* 48, 107–133.

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: Bewley, Truman (Ed.) Advances in Economic Theory: Fifth World Congress, Chapter 2, 33-70, Cambridge University Press, Cambridge, UK.

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 mechanisms that maximize traders' total expected surplus in the set of equilibrium-incentive-compatible or *L1*-incentive-compatible mechanisms with a homogenous population of *L1* traders and observable levels[21]



Equilibrium: 0.25, 0.75

Buyer (—) and seller (···)

L1: 0.25, 0.75

Equilibrium: 0.25, 1.25

Buyer (—) and seller (···)

L1: 0.25, 1.25

Equilibrium: 0.25, 1.5

Buyer (—) and seller (···)

L1: 0.25, 1.5

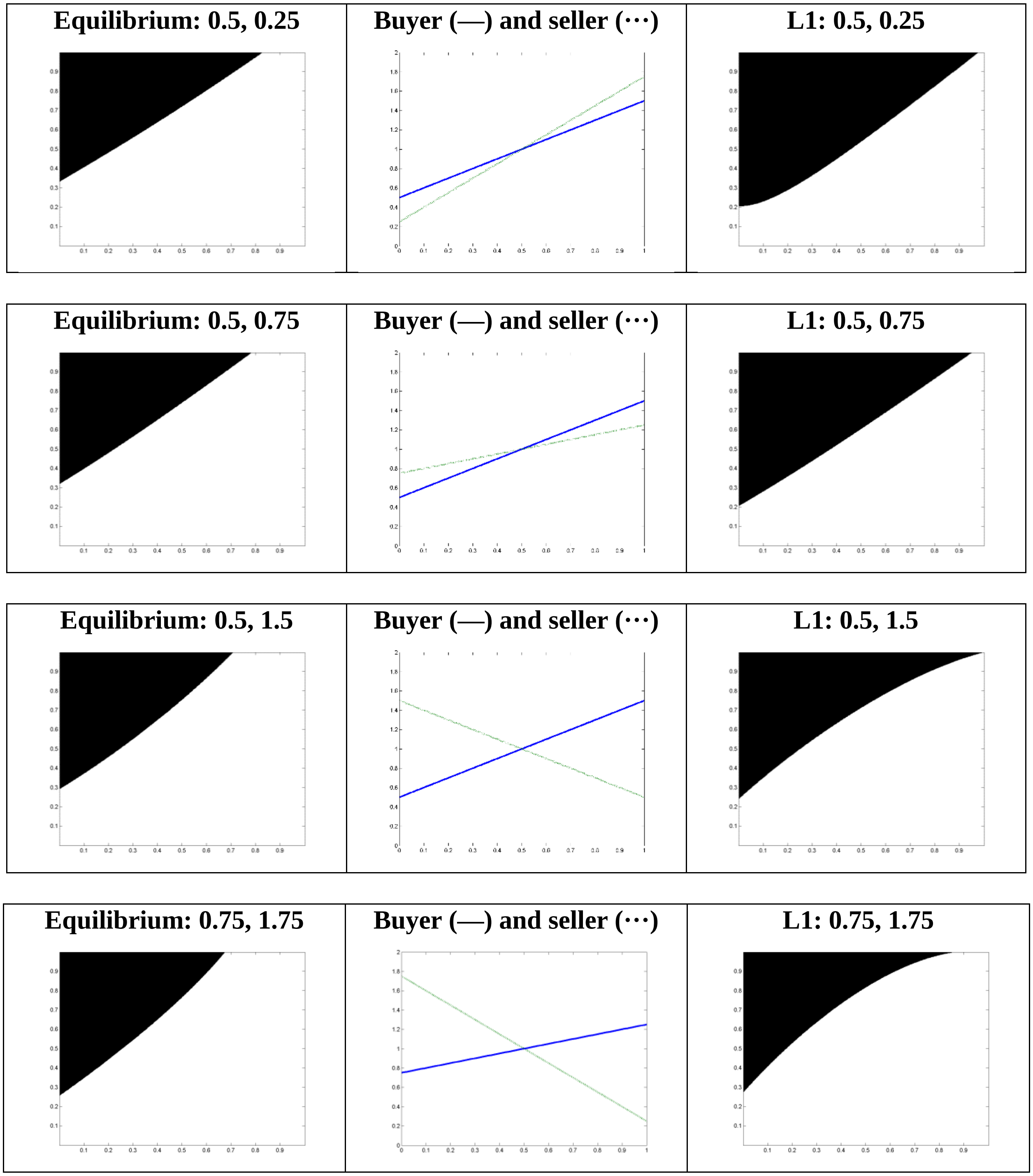
Equilibrium: 0.25, 1.75

Buyer (—) and seller (···)

L1: 0.25, 1.75

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

Figure 1 (continued). Trading regions (in black) for mechanisms that maximize traders' total expected surplus in the set of equilibrium-incentive-compatible or *L1*-incentive-compatible mechanisms with a homogenous population of *L1* traders and observable levels



Equilibrium: 0.5, 0.25

Buyer (—) and seller (···)

L1: 0.5, 0.25

Equilibrium: 0.5, 0.75

Buyer (—) and seller (···)

L1: 0.5, 0.75

Equilibrium: 0.5, 1.5

Buyer (—) and seller (···)

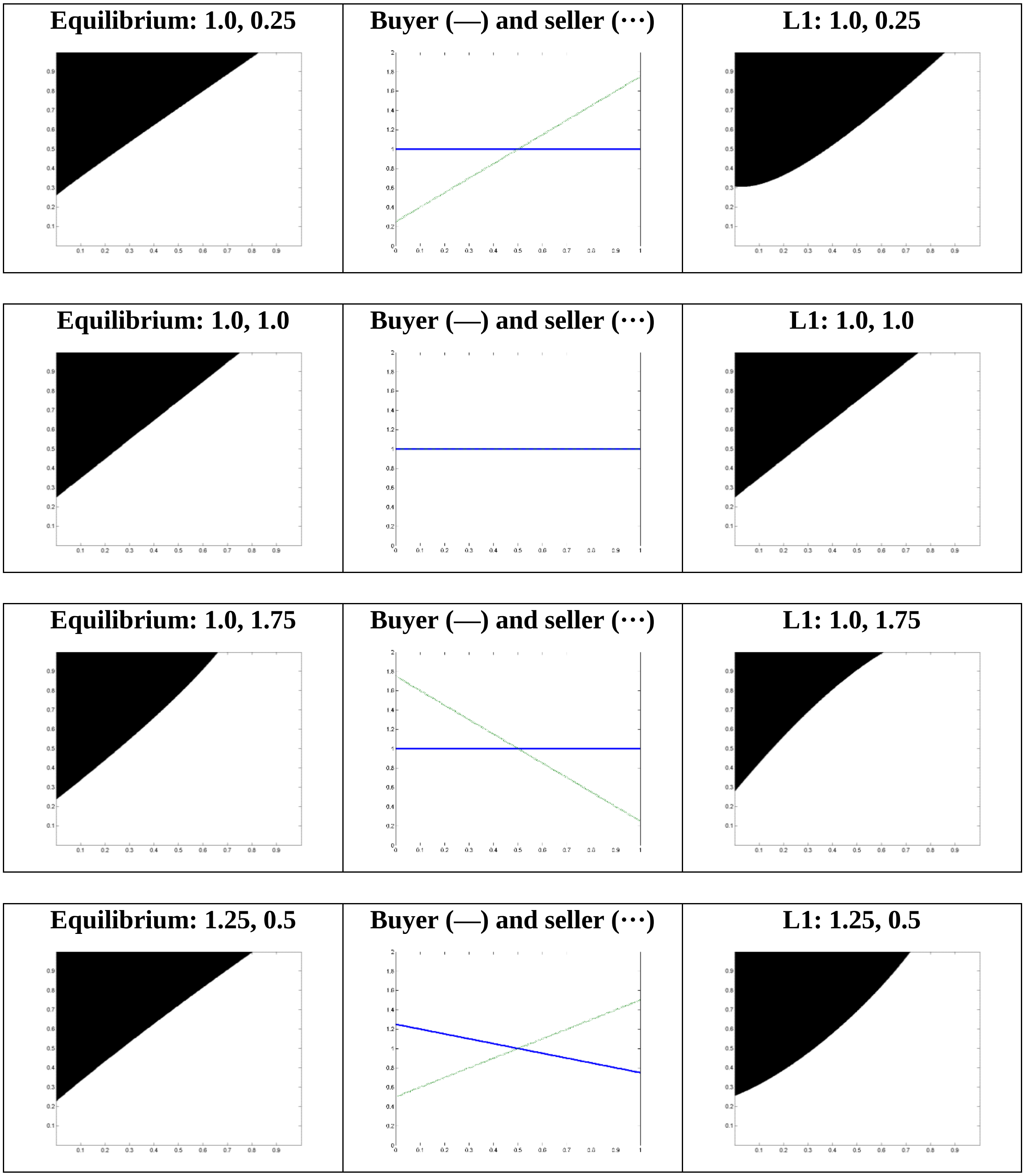
L1: 0.5, 1.5

Equilibrium: 0.75, 1.75

Buyer (—) and seller (···)

L1: 0.75, 1.75

Figure 1 (continued). Trading regions (in black) for mechanisms that maximize traders' total expected surplus in the set of equilibrium-incentive-compatible or *L1*-incentive-compatible mechanisms with a homogenous population of *L1* traders and observable levels



Equilibrium: 1.0, 0.25

Buyer (—) and seller (···)

L1: 1.0, 0.25

Equilibrium: 1.0, 1.0

Buyer (—) and seller (···)

L1: 1.0, 1.0

Equilibrium: 1.0, 1.75

Buyer (—) and seller (···)

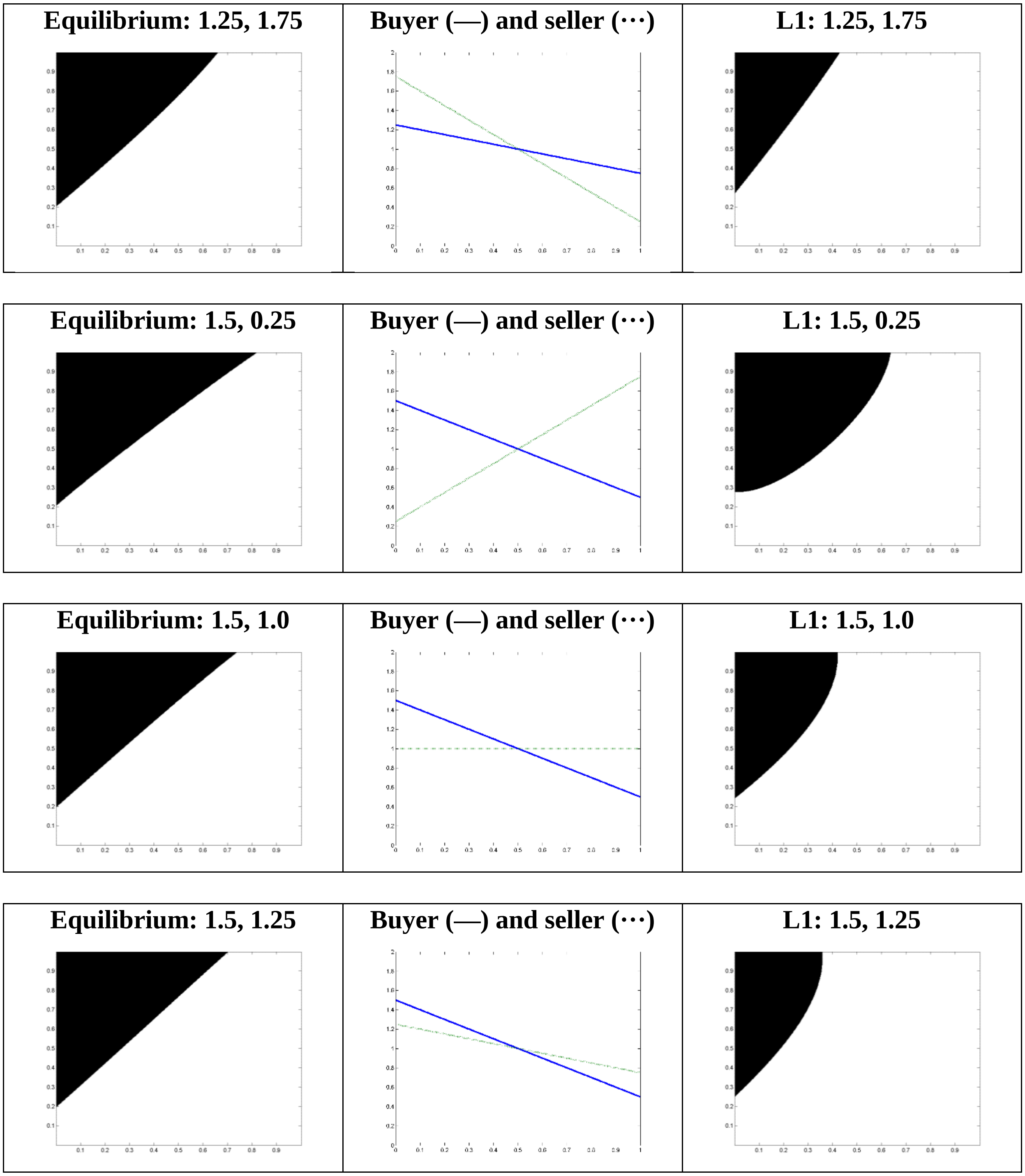
L1: 1.0, 1.75

Equilibrium: 1.25, 0.5

Buyer (—) and seller (···)

L1: 1.25, 0.5

Figure 1 (concluded). Trading regions (in black) for mechanisms that maximize traders' total expected surplus in the set of equilibrium-incentive-compatible or *L1*-incentive-compatible mechanisms with a homogenous population of *L1* traders and observable levels



Equilibrium: 1.25, 1.75

Buyer (—) and seller (···)

L1: 1.25, 1.75

Equilibrium: 1.5, 0.25

Buyer (—) and seller (···)

L1: 1.5, 0.25

Equilibrium: 1.5, 1.0

Buyer (—) and seller (···)

L1: 1.5, 1.0

Equilibrium: 1.5, 1.25

Buyer (—) and seller (···)

L1: 1.5, 1.25